



1	TTAATTAAAGGAGGACCATC	ATG AAC GAG GCC ATC GCC GTC GTC GGC ATG TCC TGC CGC CTG CCG	64
1	M N E A I A V V G M S C R L P	15	
65	AAG GCC TCG AAC CCG GCC GCC TTC TGG GAG CTG CTG CCG AAC GGG GAG AGC GCC GTC ACC	124	
16	K A S N P A A F W E L L R N G E S A V T	35	
125	GAC GTG CCC TCC GGC CCG TGG ACG TCG GTG CTC GGG GGA GCG GAC GCC GAG GAG CCG GCG	184	
36	D V P S G R W T S V L G G A D A E E P A	55	
185	GAG TCC GGT GTC CGC CCG GGC GGC TTC CTC GAC TCC CTC GAC CTC TTC GAC GCG GCC TTC	244	
56	E S G V R R G G F L D S L D L F D A A F	75	
245	TTC GGA ATC TCG CCC CGT GAG GCC GCC GCC ATG GAC CCG CAG CAG CGA CTG GTC CTC GAA	304	
76	F G I S P R E A A A M D P Q Q R L V L E	95	
305	CTC GCC TGG GAG GCG CTG GAG GAC GCC GGA ATC GTC CCC GGC ACC CTC GCC GGA AGC CCG	364	
96	L A W E A L E D A G I V P G T L A G S R	115	
365	ACC GCC GTC TTC GTC GGC ACC CTG CCG GAC GAC TAC ACG AGC CTC CTC TAC CAG CAC GGC	424	
116	T A V F V G T L R D D Y T S L L Y Q H G	135	
425	GAG CAG GCC ATC ACC CAG CAC ACC ATG GCG GGC GTG AAC CCG GGC GTC ATC GCC AAC CCG	484	
136	E Q A I T Q H T M A G V N R G V I A N R	155	
485	GTC TCG TAC CAC CTC GGC CTG CAG GGC CCG AGC CTC ACC GTC GAC GCC GCG CAG TCG TCC	544	
156	V S Y H L G L Q G P S L T V D A A Q S S	175	
545	TCG CTC GTC GCC GTG CAC CTG GCC TGC GAG TCC CTG CCG GCC GGG GAG TCC ACG ACG GCG	604	
176	S L V A V H L A C E S L R A G E S T T A	195	
605	CTC GTC GCC GGC GTG AAC CTC AAC ATC CTC GCG GAG AGC GCC GTG ACG GAG GAG CCG TTC	664	
196	L V A G V N L N I L A E S A V T E E R F	215	
665	GGT GGA CTC TCC CCG GAC GGC ACC GCC TAC ACC TTC GAC GCG CCG GCC AAC GGA TTC GTC	724	
216	G G L S P D G T A Y T F D A R A N G F V	235	
725	CGG GGC GAG GGC GGC GGA GTC GTC GTA CTC AAG CCG CTC TCC CCG GCC CTC GCC GAC GGC	784	
236	R G E G G G V V V L K P L S R A L A D G	255	
785	GAC CGT GTC CAC GGC GTC ATC CCG GCC AGC GCC GTC AAC AAC GAC GGA GCC ACC CCG GGT	844	
256	D R V H G V I R A S A V N N D G A T P G	275	
845	CTC ACC GTG CCC AGC AGG GCC GCC CAG GAG AAG GTG CTG CCG GAG GCG TAC CCG AAG GCG	904	
276	L T V P S R A A Q E K V L R E A Y R K A	295	
905	GCC CTG GAC CCG TCC GCC GTC CAG TAC GTC GAA CTC CAC GGC ACC GGA ACC CCC GTC GGC	964	
296	A L D P S A V Q Y V E L H G T G T P V G	315	
965	GAC CCC ATC GAG GCC GCC GCG CTC GGC GCC GTC CTC GGC TCG GCG CCG CCC GCG GAC GAA	1024	
316	D P I E A A A L G A V L G S A R P A D E	335	
1025	CCC CTG CTC GTC GGC TCG GCC AAG ACG AAC GTC GGG CAC CTC GAA GGC GCC GCC GGC ATC	1084	
336	P L L V G S A K T N V G H L E G A A G I	355	
1085	GTC GGC CTC ATC AAG ACG CTC CTC GCG CTC GGC CCG CCG CCG ATC CCG GCG AGC CTC AAC	1144	
356	V G L I K T L L A L G R R R I P A S L N	375	
1145	TTC CGT ACG CCC CAC CCG GAC ATC CCG CTC GAC ACC CTC GGG CTC GAC GTG CCC GAC GGC	1204	
376	F R T P H P D I P L D T L G L D V P D G	395	
1205	CTG CCG GAG TGG CCG CAC CCG GAC CCG GAA CTC CTC GCC GGC GTC AGC TCG TTC GGC ATG	1264	
396	L R E W P H P D R E L L A G V S S F G M	415	
1265	GGC GGC ACC AAC GCC CAC GTC GTC CTC AGC GAA GGC CCC GCC CAG GGC GGC GAG CAG CCC	1324	
416	G G T N A H V V L S E G P A Q G G E Q P	435	
1325	GGC ATC GAT GAG GAG ACC CCC GTC GAC AGC GCG GCC GCA CTG CCC TTC GTC GTC ACC GGC	1384	
436	G I D E E T P V D S G A A L P F V V T G	455	
1385	CGC GGC GGC GAG GCC CTG CCG GCC CAG GCC CCG CCG CTG CAC GAG GCC GTC GAA GCG GAC	1444	
456	R G G E A L R A Q A R R L H E A V E A D	475	

FIG. 23A

1445	COG	GAG	CTC	GCG	CCC	GCC	GCA	CTC	GCC	CGG	TOG	CTG	GTC	ACC	ACC	CGT	ACG	GTC	TTC	ACG	1504
476	P	E	L	A	P	A	A	L	A	R	S	L	V	T	T	R	T	V	F	T	495
1505	CAC	CGG	TOG	GTC	GTC	CTC	GCC	CGG	GAC	CGC	GCC	CGC	CTC	CTC	GAC	GGC	CTC	GGC	GCC	CTC	1564
496	H	R	S	V	V	L	A	P	D	R	A	R	L	L	D	G	L	G	A	L	515
1565	GCC	GCC	GGG	ACG	CCC	GCG	CCC	GGC	GTG	GTC	ACC	GGC	ACC	CCC	GCC	CCC	GGG	CGC	CTC	GCC	1624
516	A	A	G	T	P	A	P	G	V	V	T	G	T	P	A	P	G	R	L	A	535
1625	GTC	CTG	TTC	AGC	GGC	CAG	GGT	GCC	CAA	CGT	ACG	GGC	ATG	GGC	ATG	GAG	TTC	TAC	GCC	GCC	1684
536	V	L	F	S	G	Q	G	A	Q	R	T	G	M	G	M	E	L	Y	A	A	555
1685	CAC	CCC	GCC	TTC	GCG	ACG	GCC	TTC	GAC	GCC	GTC	GCC	GCC	GAA	CTG	GAC	CCC	CTC	CTC	GAC	1744
556	H	P	A	F	A	T	A	F	D	A	V	A	A	E	L	D	P	L	L	D	575
1745	CGG	CCC	CTC	GCC	GAA	CTC	GTC	GCG	GCG	GGC	GAC	ACC	CTC	GAC	CGC	ACC	GTC	CAC	ACA	CAG	1804
576	R	P	L	A	E	L	V	A	A	G	D	T	L	D	R	T	V	H	T	Q	595
1805	CCC	GCG	CTC	TTC	GCC	GTG	GAG	GTC	GCC	CTC	CAC	CGC	CTC	GTC	GAG	TCC	TGG	GGC	GTC	ACG	1864
596	P	A	L	F	A	V	E	V	A	L	H	R	L	V	E	S	W	G	V	T	615
1865	CCC	GAC	CTG	CTC	GCC	GGC	CAC	TCC	GTC	GGC	GAG	ATC	AGC	GCC	GCC	CAC	GTC	GCC	GGG	GTC	1924
616	P	D	L	L	A	G	H	S	V	G	E	I	S	A	A	H	V	A	G	V	635
1925	CTG	TOG	CTG	CGC	GAC	GCC	GCC	CGC	CTC	GTC	GCG	GCG	CGC	GGC	CGC	CTC	ATG	CAG	GCG	CTC	1984
636	L	S	L	R	D	A	A	R	L	V	A	A	R	G	R	L	M	Q	A	L	655
1985	CCC	GAG	GGC	GGC	GCG	ATG	GTC	GCG	GTC	GAG	GCG	AGC	GAG	GAG	GAA	GTG	CTT	CCG	CAC	CTC	2044
656	P	E	G	G	A	M	V	A	V	E	A	S	E	E	E	V	L	P	H	L	675
2045	GCG	GGA	CGC	GAG	CGG	GAG	CTC	TCC	CTC	GCG	GCC	GTG	AAC	GGC	CCC	CGC	GCG	GTC	GTC	CTC	2104
676	A	G	R	E	R	E	L	S	L	A	A	V	N	G	P	R	A	V	V	L	695
2105	GCG	GGC	GCC	GAG	CGC	GCC	GTC	CTC	GAC	GTC	GCC	GAG	CTG	CTG	CGC	GAA	CAG	GGC	CGC	CGG	2164
696	A	G	A	E	R	A	V	L	D	V	A	E	L	L	R	E	Q	G	R	R	715
2165	ACG	AAG	CGG	CTC	AGC	GTC	TOG	CAC	GCC	TTC	CAC	TOG	CCG	CTC	ATG	GAG	CCG	ATG	CTC	GAC	2224
716	T	K	R	L	S	V	S	H	A	F	H	S	P	L	M	E	P	M	L	D	735
2225	GAC	TTC	CGC	CGG	GTC	GTC	GAA	GAG	CTG	GAC	TTC	CAG	GAG	CCC	CGC	GTC	GAC	GTC	GTG	TCC	2284
736	D	F	R	R	V	V	E	E	L	D	F	Q	E	P	R	V	D	V	V	S	755
2285	ACG	GTG	ACG	GGC	CTG	CCT	GTC	ACA	GCG	GGC	CAA	TGG	ACC	GAT	CCC	GAG	TAC	TGG	GTG	GAC	2344
756	T	V	T	G	L	P	V	T	A	G	Q	W	T	D	P	E	Y	W	V	D	775
2345	CAG	GTC	CGC	AGG	CCC	GTA	CGC	TTC	CTC	GAC	GCC	GTA	CGC	ACC	CTG	GAG	GAA	TOG	GGC	GCC	2404
776	Q	V	R	R	P	V	R	F	L	D	A	V	R	T	L	E	E	S	G	A	795
2405	GAC	ACC	TTC	CTG	GAG	CTC	GGT	CCC	GAC	GGG	GTC	TGC	TCC	GCG	ATG	GCG	GCG	GAC	TCC	GTA	2464
796	D	T	F	L	E	L	G	P	D	G	V	C	S	A	M	A	A	D	S	V	815
2465	CGC	GAC	CAG	GAG	GCC	GCC	ACG	GCG	GTC	TCC	GCC	CTG	CGC	AAG	GGC	CGC	CCG	GAG	CCC	CAG	2524
816	R	D	Q	E	A	A	T	A	V	S	A	L	R	K	G	R	P	E	P	Q	835
2525	TOG	CTG	CTC	GCC	GCA	CTC	ACC	ACC	GTC	TTC	GTC	CGG	GGC	CAC	GAC	GTC	GAC	TGG	ACC	GCC	2584
836	S	L	L	A	A	L	T	T	V	F	V	R	G	H	D	V	D	W	T	A	855
2585	GCG	CAC	GGG	AGC	ACC	GGC	ACG	GTC	AGG	GTG	CCC	CTG	CCG	ACC	TAC	GCC	TTC	CAG	CGC	GAA	2644
856	A	H	G	S	T	G	T	V	R	V	P	L	P	T	Y	A	F	Q	R	E	875
2645	CGC	CAC	TGG	TTC	GAC	GGC	GCC	GCG	CGA	ACG	GCG	GCG	CCG	CTC	ACG	GCG	GGC	CGA	TOG	GGC	2704
876	R	H	W	F	D	G	A	A	R	T	A	A	P	L	T	A	G	R	S	G	895
2705	ACC	GGT	GCG	GGC	ACC	GGC	CCG	GCC	GCG	GGT	GTG	ACG	TOG	GGC	GAG	GGC	GAG	GGC	GAG	GGC	2764
896	T	G	A	G	T	G	P	A	A	G	V	T	S	G	E	G	E	G	E	G	915
2765	GAG	GGC	GCG	GGT	GCG	GGT	GGC	GGT	GAT	CGG	CCG	GCT	CGC	CAC	GAG	ACG	ACC	GAG	CGC	GTG	2824
916	E	G	A	G	A	G	G	G	D	R	P	A	R	H	E	T	T	E	R	V	935
2825	CGC	GCA	CAC	GTC	GCC	GCC	GTC	CTC	GAG	TAC	GAC	GAC	CCG	ACC	CGC	GTC	GAA	CTC	GGC	CTC	2884
936	R	A	H	V	A	A	V	L	E	Y	D	D	P	T	R	V	E	L	G	L	955
2885	ACC	TTC	AAG	GAG	CTG	GGC	TTC	GAC	TCC	CTC	ATG	TCC	GTC	GAG	CTG	CGG	AAC	GCG	CTC	GTC	2944
956	T	F	K	E	L	G	F	D	S	L	M	S	V	E	L	R	N	A	L	V	975
2945	GAC	GAC	ACG	GGA	CTG	CGC	CTG	CCC	AGC	GGA	CTG	CTC	TTC	GAC	CAC	CCG	ACG	CCG	CGC	GCC	3004
976	D	D	T	G	L	R	L	P	S	G	L	L	F	D	H	P	T	P	R	A	995

FIG. 23B

3005	CTC	GCC	GCC	CAC	CTG	GGC	GAC	CTG	CTC	ACC	GGC	GGC	AGC	GGC	GAG	ACC	GGA	TCG	GCC	GAC	3064
996	L	A	A	H	L	G	D	L	L	T	G	G	S	G	E	T	G	S	A	D	1015
3065	GGG	ATA	CCG	CCC	GCG	ACC	CCG	GCG	GAC	ACC	ACC	GCC	GAG	CCC	ATC	GCG	ATC	ATC	GGC	ATG	3124
1016	G	I	P	P	A	T	P	A	D	T	T	A	E	P	I	A	I	I	G	M	1035
3125	GCC	TGC	CGC	TAC	CCC	GGC	GGC	GTC	ACC	TCC	CCC	GAG	GAC	CTG	TGG	CGG	CTC	GTC	GCC	GAG	3184
1036	A	C	R	Y	P	G	G	V	T	S	P	E	D	L	W	R	L	V	A	E	1055
3185	GGG	CGC	GAC	GCC	GTC	TCG	GGG	CTG	CCC	ACC	GAC	CGC	GGC	TGG	GAC	GAG	GAC	CTC	TTC	GAC	3244
1056	G	R	D	A	V	S	G	L	P	T	D	R	G	W	D	E	D	L	F	D	1075
3245	GCC	GAC	CCC	GAC	CGC	AGC	GGC	AAG	AGC	TCG	GTC	CGC	GAG	GGC	GGA	TTC	CTG	CAC	GAC	GCC	3304
1076	A	D	P	D	R	S	G	K	S	S	V	R	E	G	G	F	L	H	D	A	1095
3305	GCC	CTG	TTC	GAC	GCC	GGC	TTC	TTC	GGG	ATA	TCG	CCC	CGC	GAG	GCC	CTC	GGC	ATG	GAC	CCG	3364
1096	A	L	F	D	A	G	F	F	G	I	S	P	R	E	A	L	G	M	D	P	1115
3365	CAG	CAG	CGG	CTG	CTC	CTG	GAG	ACG	GCA	TGG	GAG	GCC	GTC	GAG	CGC	GCA	GGG	CTC	GAC	CCC	3424
1116	Q	Q	R	L	L	L	E	T	A	W	E	A	V	E	R	A	G	L	D	P	1135
3425	GAA	GGC	CTC	AAG	GGC	AGC	CGG	ACG	GCC	GTC	TTC	GTC	GGC	GCC	ACC	GCC	CTG	GAC	TAC	GGC	3484
1136	E	G	L	K	G	S	R	T	A	V	F	V	G	A	T	A	L	D	Y	G	1155
3485	CCG	CGC	ATG	CAC	GAC	GGC	GCC	GAG	GGC	GTC	GAG	GGC	CAC	CTC	CTG	ACC	GGG	ACC	ACG	CCC	3544
1156	P	R	M	H	D	G	A	E	G	V	E	G	H	L	L	T	G	T	T	P	1175
3545	AGC	GTC	ATG	TCG	GGC	CGC	ATC	GCC	TAC	CAG	CTC	GGC	CTC	ACC	GGT	CCT	CGC	GTC	ACC	GTC	3604
1176	S	V	M	S	G	R	I	A	Y	Q	L	G	L	T	G	P	A	V	T	V	1195
3605	GAC	ACG	GCC	TGC	TCG	TCC	TCG	CTC	GTC	GCG	CTG	CAC	CTG	GCC	GTC	CGT	TCG	CTG	CGG	CAG	3664
1196	D	T	A	C	S	S	S	L	V	A	L	H	L	A	V	R	S	L	R	Q	1215
3665	GGC	GAG	TCG	AGC	CTC	GCG	CTC	GCC	GGC	GGA	GCG	ACC	GTC	ATG	TCG	ACA	CCG	GGC	ATG	TTC	3724
1216	G	E	S	S	L	A	L	A	G	G	A	T	V	M	S	T	P	G	M	F	1235
3725	GTC	GAG	TTC	TCG	CGG	CAG	CGC	GGC	CTC	GCC	GCC	GAC	GGC	CGC	TCC	AAG	GCC	TTC	TCC	GAC	3784
1236	V	E	F	S	R	Q	R	G	L	A	A	D	G	R	S	K	A	F	S	D	1255
3785	TCC	GCC	GAC	GGC	ACC	TCC	TGG	GCC	GAG	GGC	GTC	GGC	CTC	CTC	GTC	GTC	GAG	CGG	CTC	TCG	3844
1256	S	A	D	G	T	S	W	A	E	G	V	G	L	L	V	V	E	R	L	S	1275
3845	GAC	GCC	GAG	CGC	AAC	GGC	CAC	CCC	GTC	CTC	GCC	GTC	ATC	CGG	GGC	AGC	CGG	GTC	AAC	CAG	3904
1276	D	A	E	R	N	G	H	P	V	L	A	V	I	R	G	S	A	V	N	Q	1295
3905	GAC	GGC	GCC	TCC	AAC	GGG	CTC	ACC	GCC	CCC	AAC	GGC	CCG	TCC	CAG	CAG	CGC	GTC	ATC	CGA	3964
1296	D	G	A	S	N	G	L	T	A	P	N	G	P	S	Q	Q	R	V	I	R	1315
3965	CAG	GCC	CTG	GCC	GAC	GCC	GGG	CTC	ACC	CCG	GCC	GAC	GTC	GAC	GCC	GTC	GAG	GCG	CAC	GGT	4024
1316	Q	A	L	A	D	A	G	L	T	P	A	D	V	D	A	V	E	A	H	G	1335
4025	ACG	GGT	ACC	CGG	CTC	GGC	GAC	CCC	ATC	GAG	GCC	GAG	GCG	ATC	CTC	GGC	ACC	TAC	GGC	CGG	4084
1336	T	G	T	R	L	G	D	P	I	E	A	E	A	I	L	G	T	Y	G	R	1355
4085	GAC	CGG	GGC	GAG	GGC	GCT	CCG	CTC	CAG	CTC	GGC	TCG	CTG	AAG	TCG	AAC	ATC	GGC	CAC	GCG	4144
1356	D	R	G	E	G	A	P	L	Q	L	G	S	L	K	S	N	I	G	H	A	1375
4145	CAG	GCC	GCC	GCG	GGC	GTC	GGC	GGG	CTC	ATC	AAG	ATG	GTC	CTC	GCG	ATG	CGC	CAC	GGC	GTC	4204
1376	Q	A	A	A	G	V	G	G	L	I	K	M	V	L	A	M	R	H	G	V	1395
4205	CTG	CCC	AGG	ACG	CTC	CAC	GTC	GAC	CGG	CCC	ACC	ACC	CGC	GTC	GAC	TGG	GAG	GCC	GGC	GGC	4264
1396	L	P	R	T	L	H	V	D	R	P	T	T	R	V	D	W	E	A	G	G	1415
4265	GTC	GAG	CTC	CTC	ACC	GAG	GAG	CGG	GAG	TGG	CCG	GAG	ACG	GGC	CGC	CCG	CGC	CGC	GCG	GCG	4324
1416	V	E	L	L	T	E	E	R	E	W	P	E	T	G	R	P	R	R	A	A	1435
4325	ATC	TCC	TCC	TTC	GGC	ATC	AGC	GGC	ACC	AAC	GCC	CAC	ATC	GTC	GTC	GAA	CAG	GCC	CCG	GAA	4384
1436	I	S	S	F	G	I	S	G	T	N	A	H	I	V	V	E	Q	A	P	E	1455
4385	GCC	GGG	GAG	GCG	GCG	GTC	ACC	ACC	ACC	GCC	CCG	GAA	GCA	GGG	GAA	GCC	GGG	GAA	GCG	GCG	4444
1456	A	G	E	A	A	V	T	T	T	A	P	E	A	G	E	A	G	E	A	A	1475
4445	GAC	ACC	ACC	GCC	ACC	ACG	ACG	CCG	GCC	GCG	GTC	GGC	GTC	CCC	GAA	CCC	GTA	CGC	GCC	CCC	4504
1476	D	T	T	A	T	T	T	P	A	A	V	G	V	P	E	P	V	R	A	P	1495
4505	GTC	GTC	GTC	TCC	GCG	CGG	GAC	GCC	GCC	GCC	CTG	CGC	GCC	CAG	GCC	GTT	CGG	CTG	CGG	ACC	4564
1496	V	V	V	S	A	R	D	A	A	A	L	R	A	Q	A	V	R	L	R	T	1515

FIG. 23C

4565	TTC	CTC	GAC	GGC	CGA	COG	GAC	GTC	ACC	GTC	GCC	GAC	CTC	GGA	CGC	TOG	CTG	GCC	GCC	CGT	4624
1516	F	L	D	G	R	P	D	V	T	V	A	D	L	G	R	S	L	A	A	R	1535
4625	ACC	GCC	TTC	GAG	CAC	AAG	GCC	GCC	CTC	ACC	ACC	GCC	ACC	AGG	GAC	GAG	CTG	CTC	GCC	GGG	4684
1536	T	A	F	E	H	K	A	A	L	T	T	A	T	R	D	E	L	L	A	G	1555
4685	CTC	GAC	GCC	CTC	GGC	CGC	GGG	GAG	CAA	GCC	ACG	GGC	CTG	GTC	ACC	GGC	GAA	CCG	GCC	AGG	4744
1556	L	D	A	L	G	R	G	E	Q	A	T	G	L	V	T	G	E	P	A	R	1575
4745	GCC	GGA	CGC	ACG	GCC	TTC	CTG	TTC	ACC	GGC	CAG	GGA	GCG	CAG	CGC	GTC	GCC	ATG	GGC	GAG	4804
1576	A	G	R	T	A	F	L	F	T	G	Q	G	A	Q	R	V	A	M	G	E	1595
4805	GAA	CTG	CGC	GCC	GCG	CAC	CCC	GTG	TTC	GCC	GCC	GCC	CTC	GAC	ACC	GTG	TAC	GCG	GCC	CTC	4864
1596	E	L	R	A	A	H	P	V	F	A	A	A	L	D	T	V	Y	A	A	L	1615
4865	GAC	CGT	CAC	CTC	GAC	CGG	CCG	CTG	CGG	GAG	ATC	GTC	GCC	GCC	GGG	GAG	GAG	CTG	GAC	CTC	4924
1616	D	R	H	L	D	R	P	L	R	E	I	V	A	A	G	E	E	L	D	L	1635
4925	ACC	GCG	TAC	ACC	CAG	CCC	GCC	CTC	TTC	GCC	TTC	GAG	GTG	GCG	CTG	TTC	CGC	CTC	CTC	GAA	4984
1636	T	A	Y	T	Q	P	A	L	F	A	F	E	V	A	L	F	R	L	L	E	1655
4985	CAC	CAC	GGC	CTC	GTC	CCC	GAC	CTG	CTC	ACC	GGC	CAC	TCC	GTG	GGC	GAG	ATC	GCC	GCC	GCG	5044
1656	H	H	G	L	V	P	D	L	L	T	G	H	S	V	G	E	I	A	A	A	1675
5045	CAC	GTC	GCC	GGT	GTC	CTC	TCC	CTC	GAC	GAC	GCC	GCA	CGT	CTC	GTC	ACC	GCC	CGC	GGC	CGG	5104
1676	H	V	A	G	V	L	S	L	D	D	A	A	R	L	V	T	A	R	G	R	1695
5105	CTC	ATG	CAG	TOG	GCC	CGC	GAG	GGC	GGC	GCG	ATG	ATC	GCC	GTG	CAG	GCG	GGC	GAG	GCC	GAG	5164
1696	L	M	Q	S	A	R	E	G	G	A	M	I	A	V	Q	A	G	E	A	E	1715
5165	GTC	GTC	GAG	TCC	CTG	AAG	GGC	TAC	GAG	GGC	AGG	GTC	GCC	GTC	GCC	GCC	GTC	AAC	GGA	CCC	5224
1716	V	V	E	S	L	K	G	Y	E	G	R	V	A	V	A	A	V	N	G	P	1735
5225	ACC	GCC	GTG	GTC	GTC	TCC	GGC	GAC	GCG	GAC	GCC	GCC	GAG	GAG	ATC	CGC	GCC	GTA	TGG	GCG	5284
1736	T	A	V	V	V	S	G	D	A	D	A	A	E	E	I	R	A	V	W	A	1755
5285	GGA	CGC	GGC	CGG	CGC	ACC	CGC	AGG	CTG	CGC	GTC	AGC	CAC	GCC	TTC	CAC	TCC	CCG	CAC	ATG	5344
1756	G	R	G	R	R	T	R	R	L	R	V	S	H	A	F	H	S	P	H	M	1775
5345	GAC	GAC	GTC	CTC	GAC	GAG	TTC	CTC	CGG	GTC	GCC	GAG	GGC	CTG	ACC	TTC	GAG	GAG	CCG	CGG	5404
1776	D	D	V	L	D	E	F	L	R	V	A	E	G	L	T	F	E	E	P	R	1795
5405	ATC	CCC	GTC	GTC	TCC	ACG	GTC	ACC	GGC	GCG	CTC	GTC	ACG	TCC	GGC	GAG	CTC	ACC	TOG	CCC	5464
1796	I	P	V	V	S	T	V	T	G	A	L	V	T	S	G	E	L	T	S	P	1815
5465	GCG	TAC	TGG	GTC	GAC	CAG	ATC	CGG	CGG	CCC	GTG	CGC	TTC	CTG	GAC	GCC	GTC	CGC	ACC	CTG	5524
1816	A	Y	W	V	D	Q	I	R	R	P	V	R	F	L	D	A	V	R	T	L	1835
5525	GCC	GCC	CAG	GAC	GCG	ACC	GTC	CTC	GTC	GAG	ATC	GGC	CCC	GAC	GCC	GTC	CTC	ACG	GCA	CTC	5584
1836	A	A	Q	D	A	T	V	L	V	E	I	G	P	D	A	V	L	T	A	L	1855
5585	GCC	GAG	GAG	GCT	CTC	GCG	CCC	GGC	ACG	GAC	GCC	COG	GAC	GCC	CGG	GAC	GTC	ACG	GTC	GTC	5644
1856	A	E	E	A	L	A	P	G	T	D	A	P	D	A	R	D	V	T	V	V	1875
5645	COG	CTG	CTG	CGC	GCG	GGG	CGC	CCC	GAG	CCC	GAG	ACC	CTC	GCC	GCC	GGT	CTC	GCG	ACC	GCC	5704
1876	P	L	L	R	A	G	R	P	E	P	E	T	L	A	A	G	L	A	T	A	1895
5705	CAT	GTC	CAC	GGC	GCA	CCC	TTG	GAC	CGG	GCG	TOG	TTC	TTC	CCG	GAC	GGG	CGC	CGC	ACG	GAC	5764
1896	H	V	H	G	A	P	L	D	R	A	S	F	F	P	D	G	R	R	T	D	1915
5765	CTG	CCC	ACG	TAC	GCC	TTC	CGG	CGC	GAG	CAC	TAC	TGG	CTG	ACG	CCC	GAG	GCC	CGT	ACG	GAC	5824
1916	L	P	T	Y	A	F	R	R	E	H	Y	W	L	T	P	E	A	R	T	D	1935
5825	GCC	CGC	GCA	CTC	GGC	TTC	GAC	CCG	GCG	CGG	CAC	COG	CTG	CTG	ACG	ACC	ACG	GTC	GAG	GTC	5884
1936	A	R	A	L	G	F	D	P	A	R	H	P	L	L	T	T	T	V	E	V	1955
5885	GCC	GGC	GGC	GAC	GGC	GTC	CTG	CTG	ACC	GGC	CGT	CTC	TCC	CTG	ACC	GAC	CAG	CCC	TGG	CTG	5944
1956	A	G	G	D	G	V	L	L	T	G	R	L	S	L	T	D	Q	P	W	L	1975
5945	GCC	GAC	CAC	ATG	GTC	AAC	GGC	GCC	GTC	CTG	TTG	CCG	GCC	ACC	GCC	TTC	CTG	GAG	CTC	GCC	6004
1976	A	D	H	M	V	N	G	A	V	L	L	P	A	T	A	F	L	E	L	A	1995
6005	CTC	GCG	GCG	GGC	GAC	CAC	GTC	GGG	GCG	GTC	CGG	GTG	GAG	GAA	CTC	ACC	CTC	GAA	GCG	CCG	6064
1996	L	A	A	G	D	H	V	G	A	V	R	V	E	E	L	T	L	E	A	P	2015
6065	CTC	GTC	CTG	CCC	GAG	CGG	GGC	GCC	GTC	CGC	ATC	CAG	GTC	GGC	GTG	AGC	GGC	GAC	GGC	GAG	6124
2016	L	V	L	P	E	R	G	A	V	R	I	Q	V	G	V	S	G	D	G	E	2035

FIG. 23D

6125	TCG	CCG	GCC	GGG	CGC	ACC	TTC	GGT	GTG	TAC	AGC	ACC	CCC	GAC	TCC	GGC	GAC	ACC	GGT	GAC	6184
2036	S	P	A	G	R	T	F	G	V	Y	S	T	P	D	S	G	D	T	G	D	2055
6185	GAC	GCG	CCC	CGG	GAG	TGG	ACC	CGC	CAT	GTC	TCC	GGC	GTA	CTC	GGC	GAA	GGG	GAC	CCG	GCC	6244
2056	D	A	P	R	E	W	T	R	H	V	S	G	V	L	G	E	G	D	P	A	2075
6245	ACG	GAG	TCG	GAC	CAC	CCC	GGC	ACC	GAC	GGG	GAC	GGT	TCA	GCG	GCC	TGG	CCG	CCT	GCG	GCG	6304
2076	T	E	S	D	H	P	G	T	D	G	D	G	S	A	A	W	P	P	A	A	2095
6305	GCG	ACC	GCC	ACA	CCC	CTC	GAC	GGC	GTC	TAC	GAC	CGG	CTC	GCG	GAG	CTC	GGC	TAC	GGA	TAC	6364
2096	A	T	A	T	P	L	D	G	V	Y	D	R	L	A	E	L	G	Y	G	Y	2115
6365	GGT	CCG	GCC	TTC	CAG	GGC	CTG	ACG	GGG	CTG	TGG	CGC	GAC	GGC	GCC	GAC	ACG	CTC	GCC	GAG	6424
2116	G	P	A	F	Q	G	L	T	G	L	W	R	D	G	A	D	T	L	A	E	2135
6425	ATC	CGG	CTG	CCC	GCG	GCG	CAG	CAC	GAG	AGC	GCG	GGG	CTC	TTC	GGC	GTA	CAC	CCG	GCG	CTG	6484
2136	I	R	L	P	A	A	Q	H	E	S	A	G	L	F	G	V	H	P	A	L	2155
6485	CTC	GAC	GCG	GCG	CTC	CAC	CCG	ATC	GTC	CTG	GAG	GGC	AAC	TCA	GCT	GCC	GGT	GCC	TGT	GAC	6544
2156	L	D	A	A	L	H	P	I	V	L	E	G	N	S	A	A	G	A	C	D	2175
6545	GCC	GAT	ACC	GAC	GCG	ACC	GAC	CGG	ATC	CGG	CTG	CGG	TTC	GCG	TGG	GCG	GGG	GTG	ACC	CTC	6604
2176	A	D	T	D	A	T	D	R	I	R	L	P	F	A	W	A	G	V	T	L	2195
6605	CAC	GCC	GAA	GGG	GCC	ACC	GCG	CTC	CGC	GTA	CGG	ATC	ACA	CCC	ACC	GGC	CCG	GAC	ACG	GTC	6664
2196	H	A	E	G	A	T	A	L	R	V	R	I	T	P	T	G	P	D	T	V	2215
6665	ACG	CTC	CGC	CTC	ACC	GAC	ACC	ACC	GGT	GCG	CCC	GTG	GCC	ACC	GTG	GAG	TCC	CTG	ACC	CTG	6724
2216	T	L	R	L	T	D	T	T	G	A	P	V	A	T	V	E	S	L	T	L	2235
6725	CGC	GCG	GTG	GCG	AAG	GAC	CGG	CTG	GGC	ACC	ACC	GCC	GGG	CGC	GTC	GAC	GAC	GCC	CTG	TTC	6784
2236	R	A	V	A	K	D	R	L	G	T	T	A	G	R	V	D	D	A	L	F	2255
6785	ACG	GTC	GTG	TGG	ACG	GAG	ACC	GGC	ACA	CCG	GAA	CCC	GCA	GGG	CGC	GGA	GCC	GTG	GAG	GTC	6844
2256	T	V	V	W	T	E	T	G	T	P	E	P	A	G	R	G	A	V	E	V	2275
6845	GAG	GAA	CTC	GTC	GAC	CTC	GCC	GGC	CTC	GGC	GAC	CTC	GTG	GAG	CTC	GGC	GCC	GCG	GAC	GTC	6904
2276	E	E	L	V	D	L	A	G	L	G	D	L	V	E	L	G	A	A	D	V	2295
6905	GTC	CTC	CGG	GCC	GAC	CGC	TGG	ACG	CTC	GAC	GGG	GAC	CCG	TCC	GCC	GCC	GCG	CGC	ACA	GCC	6964
2296	V	L	R	A	D	R	W	T	L	D	G	D	P	S	A	A	A	R	T	A	2315
6965	GTC	CGG	CGC	ACC	CTC	GCC	ATC	GTC	CAG	GAG	TTC	CTG	TCC	GAG	CCG	CGC	TTC	GAC	GGC	TCG	7024
2316	V	R	R	T	L	A	I	V	Q	E	F	L	S	E	P	R	F	D	G	S	2335
7025	CGA	CTG	GTG	TGC	GTC	ACC	AGG	GGC	GCG	GTC	GCC	GCA	CTC	CCC	GGC	GAG	GAC	GTC	ACC	TCC	7084
2336	R	L	V	C	V	T	R	G	A	V	A	A	L	P	G	E	D	V	T	S	2355
7085	CTC	GCC	ACC	GGC	CCC	CTC	TGG	GGC	CTC	GTC	CGC	TCC	GCC	CAG	TCC	GAG	AAC	CCG	GGA	CGC	7144
2356	L	A	T	G	P	L	W	G	L	V	R	S	A	Q	S	E	N	P	G	R	2375
7145	CTG	TTC	CTC	CTG	GAC	CTG	GGT	GAA	GGC	GAA	GGC	GAG	CGC	GAC	GGA	GCC	GAG	GAG	CTG	ATC	7204
2376	L	F	L	L	D	L	G	E	G	E	G	E	R	D	G	A	E	E	L	I	2395
7205	CGC	GCG	GCC	ACG	GCC	GGG	GAC	GAG	CCG	CAG	CTC	GCG	GCA	CGG	GAC	GGC	CGA	CTG	CTC	GCG	7264
2396	R	A	A	T	A	G	D	E	P	Q	L	A	A	R	D	G	R	L	L	A	2415
7265	CCG	AGG	CTG	GCC	CGT	ACC	GCC	GCC	CTT	TCG	AGT	GAG	GAC	ACC	GCC	GGC	GGC	GCC	GAC	CGT	7324
2416	P	R	L	A	R	T	A	A	L	S	S	E	D	T	A	G	G	A	D	R	2435
7325	TTC	GGC	CCC	GAC	GGC	ACC	GTC	CTC	GTC	ACC	GGG	GGC	ACC	GGA	GGC	CTC	GGA	GCG	CTC	CTC	7384
2436	F	G	P	D	G	T	V	L	V	T	G	G	T	G	G	L	G	A	L	L	2455
7385	GCC	CGC	CAC	CTC	GTG	GAG	CGT	CAC	GGG	GTG	CGC	CGG	CTG	CTG	CTG	GTG	AGC	CGC	CGC	GGG	7444
2456	A	R	H	L	V	E	R	H	G	V	R	R	L	L	L	V	S	R	R	G	2475
7445	GCC	GAC	GCC	CCC	GGC	GCG	GCC	GAC	CTG	GGC	GAG	GAC	CTC	GCG	GGC	CTC	GGC	GCG	GAG	GTG	7504
2476	A	D	A	P	G	A	A	D	L	G	E	D	L	A	G	L	G	A	E	V	2495
7505	GCG	TTC	GCC	GCC	GCC	GAC	GCC	GCC	GAC	CGC	GAG	AGC	CTG	GCG	CGG	GCG	ATC	GCC	ACC	GTG	7564
2496	A	F	A	A	A	D	A	A	D	R	E	S	L	A	R	A	I	A	T	V	2515
7565	CCC	GCC	GAG	CAT	CCG	CTG	ACG	GCC	GTC	GTG	CAC	ACG	GCG	GGA	GTC	GTC	GAC	GAC	GCG	ACG	7624
2516	P	A	E	H	P	L	T	A	V	V	H	T	A	G	V	V	D	A	T		2535
7625	GTG	GAG	GCG	CTC	ACA	CCG	GAA	CGG	CTG	GAC	GCG	GTA	CTG	CGC	CCG	AAG	GTC	GAC	GCC	GCG	7684
2536	V	E	A	L	T	P	E	R	L	D	A	V	L	R	P	K	V	D	A	A	2555

FIG. 23E

7685	TGG	AAC	CTG	CAC	GAG	CTC	ACC	AAG	GAC	CTG	CGG	CTC	GAC	GCC	TTC	GTC	CTC	TTC	TCC	TCC	7744
2556	W	N	L	H	E	L	T	K	D	L	R	L	D	A	F	V	L	F	S	S	2575
7745	GTC	TCC	GGC	ATC	GTC	GGC	ACC	GCC	GGC	CAG	GCC	AAC	TAC	GCG	GCG	GCC	AAC	ACG	GGC	CTC	7804
2576	V	S	G	I	V	G	T	A	G	Q	A	N	Y	A	A	A	N	T	G	L	2595
7805	GAC	GCC	CTC	GCC	GCC	CAC	CGC	GCC	GCC	ACG	GGC	CTG	GCC	GCC	ACG	TCG	CTG	GCC	TGG	GGC	7864
2596	D	A	L	A	A	H	R	A	A	T	G	L	A	A	T	S	L	A	W	G	2615
7865	CTC	TGG	GAC	GGC	ACG	CAC	GGC	ATG	GGC	GGC	ACG	CTC	GGC	GCC	GCC	GAC	CTC	GCC	CGC	TGG	7924
2616	L	W	D	G	T	H	G	M	G	G	T	L	G	A	A	D	L	A	R	W	2635
7925	AGC	CGG	GCC	GGA	ATC	ACC	CCG	CTC	ACC	CCG	CTG	CAG	GGC	CTC	GCG	CTC	TTC	GAC	GCC	GCG	7984
2636	S	R	A	G	I	T	P	L	T	P	L	Q	G	L	A	L	F	D	A	A	2655
7985	GTC	GCC	AGG	GAC	GAC	GCC	CTC	CTC	GTA	CCC	GCC	GGG	CTC	CGT	CCC	ACC	GCC	CAC	CGG	GGC	8044
2656	V	A	R	D	D	A	L	L	V	P	A	G	L	R	P	T	A	H	R	G	2675
8045	ACG	GAC	GGA	CAG	CCT	CCT	GCG	CTG	TGG	CGC	GGC	CTC	GTC	CGG	GCG	CGC	CCG	CGC	CGT	GCC	8104
2676	T	D	G	Q	P	P	A	L	W	R	G	L	V	R	A	R	P	R	R	A	2695
8105	GCG	CGG	ACG	GCC	GCC	GAG	GCG	GCG	GAC	ACG	ACC	GGC	GGC	TGG	CTG	AGC	GGG	CTC	GCC	GCA	8164
2696	A	R	T	A	A	E	A	A	D	T	T	G	G	W	L	S	G	L	A	A	2715
8165	CAG	TCC	CCC	GAG	GAG	CGG	CGC	AGC	ACA	GCC	GTC	ACG	CTC	GTG	ACG	GGT	GTC	GTC	GCG	GAC	8224
2716	Q	S	P	E	E	R	R	S	T	A	V	T	L	V	T	G	V	V	A	D	2735
8225	GTC	CTC	GGG	CAC	GCC	GAC	TCC	GCC	GCG	GTC	GGG	GCG	GAG	CGG	TCC	TTC	AAG	GAC	CTC	GGC	8284
2736	V	L	G	H	A	D	S	A	A	V	G	A	E	R	S	F	K	D	L	G	2755
8285	TTC	GAC	TCC	CTG	GCC	GGG	GTG	GAG	CTC	CGC	AAC	CGG	CTG	AAC	GCC	GCC	ACC	GGC	CTG	CGG	8344
2756	F	D	S	L	A	G	V	E	L	R	N	R	L	N	A	A	T	G	L	R	2775
8345	CTC	CCC	GCG	ACC	ACG	GTC	TTC	GAC	CAT	CCC	TOG	CCG	GCC	GCG	CTC	GCG	TCC	CAT	CTC	CTC	8404
2776	L	P	A	T	T	V	F	D	H	P	S	P	A	A	L	A	S	H	L	L	2795
8405	GCC	CAG	GTG	CCC	GGG	TTG	AAG	GAG	GGG	ACG	GCG	GCG	ACC	GCG	ACC	GTC	GTG	GCC	GAG	CGG	8464
2796	A	Q	V	P	G	L	K	E	G	T	A	A	T	A	T	V	V	A	E	R	2815
8465	GGC	GCT	TCC	TTC	GGT	GAC	CGT	GCG	ACC	GAC	GAC	GAT	CCG	ATC	GCG	ATC	GTG	GGC	ATG	GCA	8524
2816	G	A	S	F	G	D	R	A	T	D	D	D	P	I	A	I	V	G	M	A	2835
8525	TGC	CGC	TAT	CCG	GGT	GGT	GTG	TCG	TOG	CCG	GAG	GAC	CTG	TGG	CGG	CTG	GTG	GCC	GAG	GGG	8584
2836	C	R	Y	P	G	G	V	S	S	P	E	D	L	W	R	L	V	A	E	G	2855
8585	ACG	GAC	GCG	ATC	AGC	GAG	TTC	CCC	GTC	AAC	CGC	GGC	TGG	GAC	CTG	GAG	AGC	CTC	TAC	GAC	8644
2856	T	D	A	I	S	E	F	P	V	N	R	G	W	D	L	E	S	L	Y	D	2875
8645	CCG	GAT	CCC	GAG	TCG	AAG	GGC	ACC	ACG	TAC	TGC	CGG	GAG	GGC	GGG	TTC	CTG	GAA	GGC	GCC	8704
2876	P	D	P	E	S	K	G	T	T	Y	C	R	E	G	G	F	L	E	G	A	2895
8705	GGT	GAC	TTC	GAC	GCC	GCC	TTC	TTC	GGC	ATC	TOG	CCG	CGC	GAG	GCC	CTG	GTG	ATG	GAC	CCG	8764
2896	G	D	F	D	A	A	F	F	G	I	S	P	R	E	A	L	V	M	D	P	2915
8765	CAG	CAG	CGG	CTG	CTG	CTG	GAG	GTG	TCC	TGG	GAG	GCG	CTG	GAA	CGC	GCG	GGC	ATC	GAC	CCG	8824
2916	Q	Q	R	L	L	L	E	V	S	W	E	A	L	E	R	A	G	I	D	P	2935
8825	TCC	TCG	CTG	CGC	GGC	AGC	CGC	GGT	GGT	GTC	TAC	GTG	GGC	GCC	GCG	CAC	GGC	TCG	TAC	GCC	8884
2936	S	S	L	R	G	S	R	G	G	V	Y	V	G	A	A	H	G	S	Y	A	2955
8885	TCC	GAT	CCC	CGG	CTG	GTG	CCC	GAG	GGC	TCG	GAG	GGC	TAT	CTG	CTG	ACC	GGC	AGC	GCC	GAC	8944
2956	S	D	P	R	L	V	P	E	G	S	E	G	Y	L	L	T	G	S	A	D	2975
8945	GCG	GTG	ATG	TCC	GGC	CGC	ATC	TCC	TAC	GCG	CTC	GGT	CTC	GAA	GGA	CCG	TCC	ATG	ACG	GTG	9004
2976	A	V	M	S	G	R	I	S	Y	A	L	G	L	E	G	P	S	M	T	V	2995
9005	GAG	ACG	GCC	TGC	TCC	TCC	TCG	CTG	GTG	GCG	CTG	CAT	CTG	GCG	GTA	CGG	GCG	CTG	CGG	CAC	9064
2996	E	T	A	C	S	S	S	L	V	A	L	H	L	A	V	R	A	L	R	H	3015
9065	GGC	GAG	TGC	GGG	CTC	GCG	CTG	GCG	GGC	GGG	GTG	GCG	GTG	ATG	GCC	GAT	CCG	GCG	GCG	TTC	9124
3016	G	E	C	G	L	A	L	A	G	G	V	A	V	M	A	D	P	A	A	F	3035
9125	GTG	GAG	TTC	TCC	CGG	CAG	AAG	GGG	CTG	GCC	GCC	GAC	GGC	CGC	TGC	AAG	GCG	TTC	TCG	GCC	9184
3036	V	E	F	S	R	Q	K	G	L	A	A	D	G	R	C	K	A	F	S	A	3055
9185	GCC	GCC	GAC	GGC	ACC	GGC	TGG	GCC	GAG	GGC	GTC	GGC	GTG	CTC	GTC	CTG	GAG	CGG	CTG	TCG	9244
3056	A	A	D	G	T	G	W	A	E	G	V	G	V	L	V	L	E	R	L	S	3075

FIG. 23F

9245	GAC	GCG	CGC	CGC	GCG	GGG	CAC	ACG	GTC	CTC	GGC	CTG	GTC	ACC	GGC	ACC	GCG	GTC	AAC	CAG	9304
3076	D	A	R	R	A	G	H	T	V	L	G	L	V	T	G	T	A	V	N	Q	3095
9305	GAC	GGT	GCC	TCC	AAC	GGG	CTG	ACC	GCG	CCC	AAC	GGC	CCA	GCC	CAG	CAA	CGC	GTC	ATC	GCC	9364
3096	D	G	A	S	N	G	L	T	A	P	N	G	P	A	Q	Q	R	V	I	A	3115
9365	GAG	GCG	CTC	GCC	GAC	GCC	GGG	CTG	TCC	CCG	GAG	GAC	GTG	GAC	GCG	GTC	GAG	GCG	CAC	GGC	9424
3116	E	A	L	A	D	A	G	L	S	P	E	D	V	D	A	V	E	A	H	G	3135
9425	ACC	GGC	ACC	CGG	CTC	GGC	GAC	CCC	ATC	GAG	GCC	GGG	GCG	CTG	CTC	GCC	GCC	TCC	GGA	CGG	9484
3136	T	G	T	R	L	G	D	P	I	E	A	G	A	L	L	A	A	S	G	R	3155
9485	AAC	CGT	TCC	GGC	GAC	CAC	CCG	CTG	TGG	CTC	GGC	TCG	CTG	AAG	TCC	AAC	ATC	GGG	CAT	GCC	9544
3156	N	R	S	G	D	H	P	L	W	L	G	S	L	K	S	N	I	G	H	A	3175
9545	CAG	GCC	GCC	GCC	GGT	GTC	GGC	GGC	GTC	ATC	AAG	ATG	CTC	CAG	GCG	CTG	CGG	CAC	GGC	TTG	9604
3176	Q	A	A	A	G	V	G	G	V	I	K	M	L	Q	A	L	R	H	G	L	3195
9605	CTG	CCC	CGC	ACC	CTC	CAC	GCC	GAC	GAG	COG	ACC	CCG	CAT	GCC	GAC	TGG	AGC	TCC	GGC	CGG	9664
3196	L	P	R	T	L	H	A	D	E	P	T	P	H	A	D	W	S	S	G	R	3215
9665	GTA	CGG	CTG	CTC	ACC	TCC	GAG	GTG	CCG	TGG	CAG	CGG	ACC	GGC	CGG	CCC	CGG	CGG	ACC	GGG	9724
3216	V	R	L	L	T	S	E	V	P	W	Q	R	T	G	R	P	R	R	T	G	3235
9725	GTG	TCC	GCC	TTC	GGC	GTC	GGC	GGC	ACC	AAT	GCC	CAT	GTC	GTC	CTC	GAA	GAG	GCA	CCC	GCC	9784
3236	V	S	A	F	G	V	G	G	T	N	A	H	V	V	L	E	E	A	P	A	3255
9785	CCG	CCC	GCG	CCG	GAA	CCG	GCC	GGG	GAG	GCC	CCC	GGC	GGC	TCC	CGC	GCC	GCA	GAA	GGG	GCG	9844
3256	P	P	A	P	E	P	A	G	E	A	P	G	G	S	R	A	A	E	G	A	3275
9845	GAA	GGG	CCC	CTG	GCC	TGG	GTG	GTC	TCC	GGA	CGC	GAC	GAG	CCG	GCC	CTG	CGG	TCC	CAG	GCC	9904
3276	E	G	P	L	A	W	V	V	S	G	R	D	E	P	A	L	R	S	Q	A	3295
9905	CGG	CGG	CTC	CGC	GAC	CAC	CTC	TCC	CGC	ACC	CCC	GGG	GCC	CGC	CCG	CGT	GAC	ATC	GCC	TTC	9964
3296	R	R	L	R	D	H	L	S	R	T	P	G	A	R	P	R	D	I	A	F	3315
9965	TCC	CTC	GCC	GCC	ACG	CGC	GCA	GCC	TTT	GAC	CAC	CGC	GCC	GTG	CTG	ATC	GGC	TCG	GAC	GGG	10024
3316	S	L	A	A	T	R	A	A	F	D	H	R	A	V	L	I	G	S	D	G	3335
10025	GCC	GAA	CTC	GCC	GCC	GCC	CTG	GAC	GCG	TTG	GCC	GAA	GGA	CGC	GAC	GGT	CCG	GCG	GTG	GTG	10084
3336	A	E	L	A	A	A	L	D	A	L	A	E	G	R	D	G	P	A	V	V	3355
10085	CGC	GGA	GTC	CGC	GAC	CGG	GAC	GGC	AGG	ATG	GCC	TTC	CTC	TTC	ACC	GGG	CAG	GGC	AGC	CAG	10144
3356	R	G	V	R	D	R	D	G	R	M	A	F	L	F	T	G	Q	G	S	Q	3375
10145	CGC	GCC	GGG	ATG	GCC	CAC	GAC	CTG	CAT	GCC	GCC	CAT	ACC	TTC	TTC	GCG	TCC	GCC	CTC	GAC	10204
3376	R	A	G	M	A	H	D	L	H	A	A	H	T	F	F	A	S	A	L	D	3395
10205	GAG	GTG	ACG	GAC	CGT	CTC	GAC	COG	CTG	CTC	GGC	CGG	COG	CTC	GGC	GCG	CTG	CTG	GAC	GCC	10264
3396	E	V	T	D	R	L	D	P	L	L	G	R	P	L	G	A	L	L	D	A	3415
10265	CGA	CCC	GGC	TCG	CCC	GAA	GCG	GCA	CTC	CTG	GAC	CGG	ACC	GAG	TAC	ACC	CAG	CCG	GCG	CTC	10324
3416	R	P	G	S	P	E	A	A	L	L	D	R	T	E	Y	T	Q	P	A	L	3435
10325	TTC	GCC	GTC	GAG	GTG	GCG	CTC	CAC	CGG	CTG	CTG	GAG	CAC	TGG	GGG	ATG	CGC	CCC	GAC	CTG	10384
3436	F	A	V	E	V	A	L	H	R	L	L	E	H	W	G	M	R	P	D	L	3455
10385	CTG	CTG	GGG	CAC	TCG	GTG	GGC	GAA	CTG	GCG	GCC	GCC	CAC	GTC	GCG	GGT	GTG	CTC	GAT	CTC	10444
3456	L	L	G	H	S	V	G	E	L	A	A	A	H	V	A	G	V	L	D	L	3475
10445	CAC	GAC	GCC	TGC	GCG	CTG	GTG	GCC	GCC	CGC	GGC	AGG	CTG	ATG	CAG	CGC	CTG	CCG	CCC	GGC	10504
3476	D	D	A	C	A	L	V	A	A	R	G	R	L	M	Q	R	L	P	P	G	3495
10505	GGC	GCG	ATG	GTC	TCC	GTG	CGG	GCC	GGC	GAG	GAC	GAG	GTC	CGC	GCA	CTG	CTG	GCC	GGC	CGC	10564
3496	G	A	M	V	S	V	R	A	G	E	D	E	V	R	A	L	L	A	G	R	3515
10565	GAG	GAC	GCC	GTC	TGC	GTC	GCC	GCG	GTG	AAC	GGC	CCC	CGG	TCG	GTG	GTG	ATC	TCC	GGC	GCG	10624
3516	E	D	A	V	C	V	A	A	V	N	G	P	R	S	V	V	I	S	G	A	3535
10625	GAG	GAA	GCG	GTG	GCC	GAG	GCG	GCG	GCG	CAG	CTC	GCC	GGA	CGA	GGC	CGC	CGC	ACC	AGG	CGG	10684
3536	E	E	A	V	A	E	A	A	A	Q	L	A	G	R	G	R	R	T	R	R	3555
10685	CTC	CGC	GTC	GCG	CAC	GCC	TTC	CAC	TCA	CCC	CTG	ATG	GAC	GGC	ATG	CTC	GCC	GGA	TTC	CGG	10744
3556	L	R	V	A	H	A	F	H	S	P	L	M	D	G	M	L	A	G	F	R	3575
10745	GAG	GTC	GCC	GCC	GGC	CTG	CGC	TAC	CGG	GAA	CCG	GAG	CTG	ACG	GTC	GTC	TCC	ACG	GTC	ACG	10804
3576	E	V	A	A	G	L	R	Y	R	E	P	E	L	T	V	V	S	T	V	T	3595

FIG. 23G

10805	GGG	CGG	CCC	GCC	CGC	CCC	GGT	GAA	CTC	ACC	GGC	CCC	GAC	TAC	TGG	GTG	GCC	CAG	GTC	CGT	10864
3596	G	R	P	A	R	P	G	E	L	T	G	P	D	Y	W	V	A	Q	V	R	3615
10865	GAG	CCC	GTG	CGC	TTC	GCG	GAC	GCG	GTC	CGC	ACG	GCA	CAC	CGC	CTC	GGA	GCC	CGC	ACC	TTC	10924
3616	E	P	V	R	F	A	D	A	V	R	T	A	H	R	L	G	A	R	T	F	3635
10925	CTG	GAG	ACC	GGC	COG	GAC	GGC	GTG	CTG	TGC	GGC	ATG	GCA	GAG	GAG	TGC	CTG	GAG	GAC	GAC	10984
3636	L	E	T	G	P	D	G	V	L	C	G	M	A	E	E	C	L	E	D	D	3655
10985	ACC	GTG	GCC	CTG	CTG	COG	GCG	ATC	CAC	AAG	CCC	GGC	ACC	GCG	CCG	CAC	GGT	CCG	GCG	GCT	11044
3656	T	V	A	L	L	P	A	I	H	K	P	G	T	A	P	H	G	P	A	A	3675
11045	CCC	GGC	GCG	CTG	COG	GCG	GCC	GCC	GCC	GCG	TAC	GGC	COG	GGC	GCC	COG	GTG	GAC	TGG	GCC	11104
3676	P	G	A	L	R	A	A	A	A	A	Y	G	R	G	A	R	V	D	W	A	3695
11105	GGG	ATG	CAC	GCC	GAC	GGC	CCC	GAG	GGG	COG	GCC	COG	COG	GTC	GAA	CTG	CCC	GTC	CAC	GCC	11164
3696	G	M	H	A	D	G	P	E	G	P	A	R	R	V	E	L	P	V	H	A	3715
11165	TTC	COG	CAC	COG	COG	TAC	TGG	CTC	GCC	COG	GGC	COG	GCG	GCG	GAC	ACC	GAC	GAC	TGG	ATG	11224
3716	F	R	H	R	R	Y	W	L	A	P	G	R	A	A	D	T	D	D	W	M	3735
11225	TAC	COG	ATC	GGC	TGG	GAC	COG	CTG	COG	GCT	GTG	ACC	GGC	GGG	GCC	COG	ACC	GCC	GGC	COG	11284
3736	Y	R	I	G	W	D	R	L	P	A	V	T	G	G	A	R	T	A	G	R	3755
11285	TGG	CTG	GTG	ATC	CAC	CCC	GAC	AGC	COG	COG	TGC	COG	GAG	CTG	TCC	GGC	CAC	GCC	GAA	COG	11344
3756	W	L	V	I	H	P	D	S	P	R	C	R	E	L	S	G	H	A	E	R	3775
11345	GCG	CTG	COG	GCC	GCG	GGC	GCG	AGC	CCC	GTA	COG	CTG	CCC	GTG	GAC	GCT	COG	GCC	GCC	GAC	11404
3776	A	L	R	A	A	G	A	S	P	V	P	L	P	V	D	A	P	A	A	D	3795
11405	COG	GCG	TCC	TTC	GCG	GCA	CTG	CTG	COG	TCC	GCC	ACC	GGA	CCT	GAC	ACA	CGA	GGT	GAC	ACA	11464
3796	R	A	S	F	A	A	L	L	R	S	A	T	G	P	D	T	R	G	D	T	3815
11465	GCC	GCG	CCC	GTG	GCC	GGT	GTG	CTG	TGC	CTG	CTG	TCC	GAG	GAG	GAT	COG	CCC	CAT	COG	CAG	11524
3816	A	A	P	V	A	G	V	L	S	L	L	S	E	E	D	R	P	H	R	Q	3835
11525	CAC	GCC	COG	GTA	CCC	GCC	GGG	GTC	CTG	GCG	ACG	CTG	TCC	CTG	ATG	CAG	GCT	ATG	GAG	GAG	11584
3836	H	A	P	V	P	A	G	V	L	A	T	L	S	L	M	Q	A	M	E	E	3855
11585	GAG	GCG	GTG	GAG	GCT	COG	GTG	TGG	TGC	GTC	TCC	COG	GCC	GCG	GTC	GCC	GCC	GCC	GAC	COG	11644
3856	E	A	V	E	A	R	V	W	C	V	S	R	A	A	V	A	A	A	D	R	3875
11645	GAA	COG	CCC	GTC	GGC	GCG	GGC	GCC	GCC	CTG	TGG	GGG	CTG	GGG	COG	GTG	GCC	GCC	CTG	GAA	11704
3876	E	R	P	V	G	A	G	A	A	L	W	G	L	G	R	V	A	A	L	E	3895
11705	COG	CCC	ACC	COG	TGG	GGC	GGT	CTC	GTG	GAC	CTG	CCC	GCC	TGG	CCC	GGT	GCG	GCG	CAC	TGG	11764
3896	R	P	T	R	W	G	G	L	V	D	L	P	A	S	P	G	A	A	H	W	3915
11765	GCG	GCC	GCC	GTG	GAA	COG	CTC	GCC	GGT	CCC	GAG	GAC	CAG	ATC	GCC	GTG	COG	GCG	TCC	GGC	11824
3916	A	A	A	V	E	R	L	A	G	P	E	D	Q	I	A	V	R	A	S	G	3935
11825	AGT	TGG	GGC	COG	COG	CTC	ACC	AGG	CTG	COG	COG	GAC	GGC	GGC	GGC	COG	ACG	GCC	GCA	CCC	11884
3936	S	W	G	R	R	L	T	R	L	P	R	D	G	G	G	R	T	A	A	P	3955
11885	GCG	TAC	COG	COG	COG	GGC	ACG	GTG	CTC	GTC	ACC	GGT	GGC	ACC	GGC	GCG	CTC	GGC	GGG	CAT	11944
3956	A	Y	R	P	R	G	T	V	L	V	T	G	G	T	G	A	L	G	G	H	3975
11945	CTC	GCC	COG	TGG	CTC	GCC	GCG	GCG	GGC	GCC	GAA	CAC	CTG	GCG	CTC	ACC	AGC	COG	COG	GGC	12004
3976	L	A	R	W	L	A	A	A	G	A	E	H	L	A	L	T	S	R	R	G	3995
12005	COG	GAC	GCG	CCC	GGC	GCC	GCC	GGA	CTC	GAG	GCC	GAA	CTC	CTC	CTC	CTG	GCC	GCC	AAG	GTG	12064
3996	P	D	A	P	G	A	A	G	L	E	A	E	L	L	L	L	G	A	K	V	4015
12065	ACG	TTC	GCC	GCC	TGC	GAC	ACC	GCC	GAC	COG	GAC	GGC	CTC	GCC	COG	GTC	CTG	COG	GCG	ATA	12124
4016	T	F	A	A	C	D	T	A	D	R	D	G	L	A	R	V	L	R	A	I	4035
12125	COG	GAG	GAC	ACC	COG	CTC	ACC	GCG	GTG	TTC	CAC	GCC	GCG	GGC	GTA	COG	CAG	GTC	ACG	COG	12184
4036	P	E	D	T	P	L	T	A	V	F	H	A	A	G	V	P	Q	V	T	P	4055
12185	CTG	TCC	CGT	ACC	TGG	CCC	GAG	CAC	TTC	GCC	GAC	GTG	TAC	GCG	GGC	AAG	GCG	GCG	GGC	GCC	12244
4056	L	S	R	T	S	P	E	H	F	A	D	V	Y	A	G	K	A	A	G	A	4075
12245	GCG	CAC	CTG	GAC	GAA	CTG	ACC	COG	GAA	CTC	GGC	GCC	GGA	CTC	GAC	GCG	TTC	GTC	CTC	TAC	12304
4076	A	H	L	D	E	L	T	R	E	L	G	A	G	L	D	A	F	V	L	Y	4095
12305	TCC	TCC	GGC	GCC	GGC	GTC	TGG	GGC	AGC	GCC	GGC	CAG	GGT	GCC	TAC	GCC	GCC	GCC	AAC	GCC	12364
4096	S	S	G	A	G	V	W	G	S	A	G	Q	G	A	Y	A	A	A	N	A	4115

FIG. 23H

12365	GCC	CTG	GAC	GCG	CTC	GCC	CGG	CGC	CGT	GCG	GCG	GAC	GGA	CTC	CCC	GCC	ACC	TCC	ATC	GCC	12424
4116	A	L	D	A	L	A	R	R	R	A	A	D	G	L	P	A	T	S	I	A	4135
12425	TGG	GGC	GTG	TGG	GGC	GGC	GGT	ATG	GGG	GCC	GAC	GAG	GCG	GGC	GCG	GAG	TAT	CTG	GGC	12484	
4136	W	G	V	W	G	G	G	M	G	A	D	E	A	G	A	E	Y	L	G	4155	
12485	CGG	CGC	GGT	ATG	CGC	CCC	ATG	GCA	CCG	GTC	TCC	GCG	CTC	CGG	GCG	ATG	GCC	ACC	GCC	ATC	12544
4156	R	R	G	M	R	P	M	A	P	V	S	A	L	R	A	M	A	T	A	I	4175
12545	GCC	TCC	GGG	GAA	CCC	TGC	CCC	ACC	GTC	ACC	CAC	ACC	GAC	TGG	GAG	CGC	TTC	GGC	GAG	GGC	12604
4176	A	S	G	E	P	C	P	T	V	T	H	T	D	W	E	R	F	G	E	G	4195
12605	TTC	ACC	GCC	TTC	CGG	CCC	AGC	CCT	CTG	ATC	GCG	GGG	CTC	GGC	ACG	CCG	GGC	GGC	GGC	CGG	12664
4196	F	T	A	F	R	P	S	P	L	I	A	G	L	G	T	P	G	G	G	R	4215
12665	GCG	GCG	GAG	ACC	CCC	GAG	GAG	GGG	AAC	GCC	ACC	GCT	GCG	GCG	GAC	CTC	ACC	GCC	CTG	CCG	12724
4216	A	A	E	T	P	E	E	G	N	A	T	A	A	A	D	L	T	A	L	P	4235
12725	CCC	GCC	GAA	CTC	CGC	ACC	GCG	CTG	CGC	GAG	CTG	GTG	CGA	GCC	CGG	ACC	GCC	GCG	GCG	CTC	12784
4236	P	A	E	L	R	T	A	L	R	E	L	V	R	A	R	T	A	A	A	L	4255
12785	GGC	CTC	GAC	GAC	CCG	GCC	GAG	GTC	GCC	GAG	GGC	GAA	CGG	TTC	CCC	GCC	ATG	GGC	TTC	GAC	12844
4256	G	L	D	D	P	A	E	V	A	E	G	E	R	F	P	A	M	G	F	D	4275
12845	TCC	CTG	GCC	ACC	GTA	CGG	CTG	CGC	CGC	GGA	CTC	GCC	TCG	GCC	ACG	GGC	CTC	GAC	CTG	CCC	12904
4276	S	L	A	T	V	R	L	R	R	G	L	A	S	A	T	G	L	D	L	P	4295
12905	CCC	GAT	CTG	CTC	TTC	GAC	CGG	GAC	ACC	CCG	GCC	GCG	CTC	GCC	GCC	CAC	CTG	GCC	GAA	CTG	12964
4296	P	D	L	L	F	D	R	D	T	P	A	A	L	A	A	H	L	A	E	L	4315
12965	CTC	GCC	ACC	GCA	CGG	GAC	CAC	GGA	CCC	GGC	GGC	CCC	GGG	ACC	GGT	GCC	GCG	CCG	GCC	GAT	13024
4316	L	A	T	A	R	D	H	G	P	G	G	P	G	T	G	A	A	P	A	D	4335
13025	GCC	GGA	AGC	GGC	CTG	CCG	GCC	CTC	TAC	CGG	GAG	GCC	GTC	CGC	ACC	GGC	CGG	GCC	GCG	GAA	13084
4336	A	G	S	G	L	P	A	L	Y	R	E	A	V	R	T	G	R	A	A	E	4355
13085	ATG	GCC	GAA	CTG	CTC	GCC	GCC	GCT	TCC	CGG	TTC	CGC	CCC	GCC	TTC	GGG	ACG	GCG	GAC	CGG	13144
4356	M	A	E	L	L	A	A	A	S	R	F	R	P	A	F	G	T	A	D	R	4375
13145	CAG	CCG	GTG	GCC	CTC	GTG	CCG	CTG	GCC	GAC	GGC	GCG	GAG	GAC	ACC	GGG	CTC	CCG	CTG	CTC	13204
4376	Q	P	V	A	L	V	P	L	A	D	G	A	E	D	T	G	L	P	L	L	4395
13205	GTG	GGC	TGC	GCC	GGG	ACG	GCG	GTG	GCC	TCC	GGC	CCG	GTG	GAG	TTC	ACC	GCC	TTC	GCC	GGA	13264
4396	V	G	C	A	G	T	A	V	A	S	G	P	V	E	F	T	A	F	A	G	4415
13265	GCG	CTG	GCG	GAC	CTC	CCG	GCG	GCG	GCC	CCG	ATG	GCC	GCG	CTG	CCG	CAG	CCC	GGC	TTC	CTG	13324
4416	A	L	A	D	L	P	A	A	A	P	M	A	A	L	P	Q	P	G	F	L	4435
13325	CCG	GGA	GAA	CGA	GTC	CCG	GCC	ACC	CCG	GAG	GCA	TTG	TTC	GAG	GCC	CAG	GCG	GAA	GCG	CTG	13384
4436	P	G	E	R	V	P	A	T	P	E	A	L	F	E	A	Q	A	E	A	L	4455
13385	CTG	CGC	TAC	GCG	GCC	GGC	CGG	CCC	TTC	GTG	CTG	CTG	GGG	CAC	TCC	GCC	GGC	GCC	AAC	ATG	13444
4456	L	R	Y	A	A	G	R	P	F	V	L	L	G	H	S	A	G	A	N	M	4475
13445	GCC	CAC	GCC	CTG	ACC	CGT	CAT	CTG	GAG	GCG	AAC	GGT	GGC	GGC	CCC	GCA	GGG	CTG	GTG	CTC	13504
4476	A	H	A	L	T	R	H	L	E	A	N	G	G	G	P	A	G	L	V	L	4495
13505	ATG	GAC	ATC	TAC	ACC	CCC	GCC	GAC	CCC	GGC	GCG	ATG	GGC	GTC	TGG	CGG	AAC	GAC	ATG	TTC	13564
4496	M	D	I	Y	T	P	A	D	P	G	A	M	G	V	W	R	N	D	M	F	4515
13565	CAG	TGC	GTC	TGG	CGG	CGC	TCG	GAC	ATC	CCC	CCG	GAC	GAC	CAC	CGC	CTC	ACG	GCC	ATG	GGC	13624
4516	Q	W	V	W	R	R	S	D	I	P	P	D	D	H	R	L	T	A	M	G	4535
13625	GCC	TAC	CAC	CGG	CTG	CTT	CTC	GAC	TGG	TCG	CCC	ACC	CCC	GTC	CGC	GCC	CCC	GTA	CTG	CAT	13684
4536	A	Y	H	R	L	L	L	D	W	S	P	T	P	V	R	A	P	V	L	H	4555
13685	CTG	CGC	GCC	GCG	GAA	CCC	ATG	GGC	GAC	TGG	CCA	CCC	GGG	GAC	ACC	GGC	TGG	CAG	TCC	CAC	13744
4556	L	R	A	A	E	P	M	G	D	W	P	P	G	D	T	G	W	Q	S	H	4575
13745	TGG	GAC	GGC	GCG	CAC	ACC	ACC	GCC	GGC	ATC	CCC	GGA	AAC	CAC	TTC	ACG	ATG	ATG	ACC	GAA	13804
4576	W	D	G	A	H	T	T	A	G	I	P	G	N	H	F	T	M	M	T	E	4595
13805	CAC	GCC	TCC	GCC	GCC	GCC	CGG	CTC	GTG	CAC	GGC	TGG	CTC	GCG	GAA	CGG	ACC	CCG	TCC	GGG	13864
4596	H	A	S	A	A	A	R	L	V	H	G	W	L	A	E	R	T	P	S	G	4615
13865	CAG	GGC	GGG	TCA	CCG	TCC	CGC	GCG	GCG	GGG	AGA	GAG	GAG	AGG	CCG	TGA	ACACGGCAGCCGGCCCC			13928	
4616	Q	G	G	S	P	S	R	A	A	G	R	E	E	R	P	*					4631

FIG. 231

13929 GACCGGCACCGCGCGCGCGCGCACACCGCCCCCGCGCGCGCACACGACCTGTCCCGCGCGCGGAOCGAGGCTCCAACCTCA 14008

14009 CCGCGCGCGCACAGTGGTTTCGCGCGCAACCAGGGAGACCCCTACGGG ATG ATC CTG CGC GCC GGC ACC GCC 14079
1 M I L R A G T A 8

14080 GAC CCG GCA CCG TAC GAG GAA GAG ATC CCC GGG TAC CGA GCT CGA ATT CTT AAT TAA GGAG 14140
9 D P A P Y E E E I P G Y R A R I L N * 27

14141 GTCGTAG ATG AGT AAC AAG AAC AAC GAT GAG CTG CAG CGG CAG GCC TCG GAA AAC ACC CTG 14201
1 M S N K N N D E L Q R Q A S E N T L 18

14202 GGG CTG AAC CCG GTC ATC GGT ATC CGC CGC AAA GAC CTG TTG AGC TCG GCA CGC ACC GTG 14261
19 G L N P V I G I R R K D L L S S A R T V 38

14262 CTG CGC CAG GCC GTG CGC CAA CCG CTG CAC AGC GCC AAG CAT GTG GCC CAC TTT GGC CTG 14321
39 L R Q A V R Q P L H S A K H V A H F G L 58

14322 GAG CTG AAG AAC GTG CTG CTG GGC AAG TCC AGC CTT GCC CCG GAA AGC GAC GAC CGT CGC 14381
59 E L K N V L L G K S S L A P E S D D R R 78

14382 TTC AAT GAC CCG GCA TGG AGC AAC AAC CCA CTT TAC CGC CGC TAC CTG CAA ACC TAT CTG 14441
79 F N D P A W S N N P L Y R R Y L Q T Y L 98

14442 GCC TGG CGC AAG GAG CTG CAG GAC TGG ATC GGC AAC AGC GAC CTG TCG CCC CAG GAC ATC 14501
99 A W R K E L Q D W I G N S D L S P Q D I 118

14502 AGC CGC GGC CAG TTC GTC ATC AAC CTG ATG ACC GAA GCC ATG GCT CCG ACC AAC ACC CTG 14561
119 S R G Q F V I N L M T E A M A P T N T L 138

14562 TCC AAC CCG GCA GCA GTC AAA CGC TTC TTC GAA ACC GGC GGC AAG AGC CTG CTC GAT GGC 14621
139 S N P A A V K R F F E T G G K S L L D G 158

14622 CTG TCC AAC CTG GCC AAG GAC CTG GTC AAC AAC GGT GGC ATG CCC AGC CAG GTG AAC ATG 14681
159 L S N L A K D L V N N G G M P S Q V N M 178

14682 GAC GCC TTC GAG GTG GGC AAG AAC CTG GGC ACC AGT GAA GGC GCC GTG GTG TAC CGC AAC 14741
179 D A F E V G K N L G T S E G A V V Y R N 198

14742 GAT GTG CTG GAG CTG ATC CAG TAC AAG CCC ATC ACC GAG CAG GTG CAT GCC CGC CCG CTG 14801
199 D V L E L I Q Y K P I T E Q V H A R P L 218

14802 CTG GTG GTG CCG CCG CAG ATC AAC AAG TTC TAC GTA TTC GAC CTG AGC CCG GAA AAG AGC 14861
219 L V V P P Q I N K F Y V F D L S P E K S 238

14862 CTG GCA CGC TAC TGC CTG CGC TCG CAG CAG CAG ACC TTC ATC ATC AGC TGG CGC AAC CCG 14921
239 L A R Y C L R S Q Q Q T F I I S W R N P 258

14922 ACC AAA GCC CAG CGC GAA TGG GGC CTG TCC ACC TAC ATC GAC GCG CTC AAG GAG GCG GTC 14981
259 T K A Q R E W G L S T Y I D A L K E A V 278

14982 GAC GCG GTG CTG GCG ATT ACC GGC AGC AAG GAC CTG AAC ATG CTC GGT GCC TGC TCC GGC 15041
279 D A V L A I T G S K D L N M L G A C S G 298

15042 GGC ATC ACC TGC ACG GCA TTG GTC GGC CAC TAT GCC GCC CTC GGC GAA AAC AAG GTC AAT 15101
299 G I T C T A L V G H Y A A L G E N K V N 318

15102 GCC CTG ACC CTG CTG GTC AGC GTG CTG GAC ACC ACC ATG GAC AAC CAG GTC GCC CTG TTC 15161
319 A L T L L V S V L D T T M D N Q V A L F 338

15162 GTC GAC GAG CAG ACT TTG GAG GCC GCC AAG CGC CAC TCC TAC CAG GCC GGT GTG CTC GAA 15221
339 V D E Q T L E A A K R H S Y Q A G V L E 358

15222 GGC AGC GAG ATG GCC AAG GTG TTC GCC TGG ATG CGC CCC AAC GAC CTG ATC TGG AAC TAC 15281
359 G S E M A K V F A W M R P N D L I W N Y 378

15282 TGG GTC AAC AAC TAC CTG CTC GGC AAC GAG CCG CCG GTG TTC GAC ATC CTG TTC TGG AAC 15341
379 W V N N Y L L G N E P P V F D I L F W N 398

15342 AAC GAC ACC ACG CGC CTG CCG GCC GCC TTC CAC GGC GAC CTG ATC GAA ATG TTC AAG AGC 15401
399 N D T T R L P A A F H G D L I E M F K S 418

15402 AAC CCG CTG ACC CGC CCG GAC GCC CTG GAG GTT TGC GGC ACT CCG ATC GAC CTG AAA CAG 15461
419 N P L T R P D A L E V C G T P I D L K Q 438

15462 GTC AAA TGC GAC ATC TAC AGC CTT GCC GGC ACC AAC GAC CAC ATC ACC CCG TGG CAG TCA 15521
439 V K C D I Y S L A G T N D H I T P W Q S 458

FIG. 23J

15522	TGC	TAC	CGC	TCG	GCG	CAC	CTG	TTC	GGC	GGC	AAG	ATC	GAG	TTC	GTG	CTG	TCC	AAC	AGC	GGC	15581
459	C	Y	R	S	A	H	L	F	G	G	K	I	E	F	V	L	S	N	S	G	478
15582	CAC	ATC	CAG	AGC	ATC	CTC	AAC	CCG	CCA	GGC	AAC	CCC	AAG	GCG	CGC	TTC	ATG	ACC	GGT	GCC	15641
479	H	I	Q	S	I	L	N	P	P	G	N	P	K	A	R	F	M	T	G	A	498
15642	GAT	CGC	CCG	GGT	GAC	CCG	GTG	GCC	TGG	CAG	GAA	AAC	GCC	ACC	AAG	CAT	GCC	GAC	TCC	TGG	15701
499	D	R	P	G	D	P	V	A	W	Q	E	N	A	T	K	H	A	D	S	W	518
15702	TGG	CTG	CAC	TGG	CAA	AGC	TGG	CTG	GGC	GAG	CGT	GCC	GGC	GAG	CTG	GAA	AAG	GCG	CCG	ACC	15761
519	W	L	H	W	Q	S	W	L	G	E	R	A	G	E	L	E	K	A	P	T	538
15762	CGC	CTG	GGC	AAC	CGT	GCC	TAT	GCC	GCT	GGC	GAG	GCA	TCC	CCG	GGC	ACC	TAC	GTT	CAC	GAG	15821
539	R	L	G	N	R	A	Y	A	A	G	E	A	S	P	G	T	Y	V	H	E	558
15822	CGT	TGA	GCTG	CAGCGCCG	TGGCCACCTG	CGGGACGCCACGGT	GTGAATTTC														15872
559	R	*																			560

FIG. 23K

41/164



plkPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890

BamHI
▼
GGATCCGACC GTGGGTGTGA ATCTCCGGGT GCTCGCCTCG TCCTGCCCGG TTACCTGTCC GCCTCCCGCT CCAGACCAGC GGGAGGCGGA CAGGGGCATG 100

SphI
▼
CCGCGCGGCG GGCTAACGGC CCGTCGCGCG TCCGTACGAC GAGCCTCGCG CGCCTGCGG GCCCTTGGTC TGCCGGACCT GTGCGCGGGG TGCGCAGGCT 200

BstXI
▼
TCGCGCGGCG GCGTGGGCGG GTATCTGCGG CTCCCGGCGA CGGCGGCGCT GTGCTCTCTCC GAGTCATAGT CCCTGCGCGC GGGCGCACCG CCCTGGGCGG 300

SphI
▼
GCATGCGGCT GCGCGGCGCG CCGGCGGCGT AACTCGGCTG GGAGGCGTGG AAAAGGGCGA TCCATTTGGT GAGCGTGAGG TCCTTTGGCA GTCCGCGGTC 400

EcoRI
Apol
▼
CGGAATTCAG TGGCGGTGCG CGAGGGAACG GTAGTCCGC TTGGGGATGT GGGGCGGAG GATCTCCGG AGGCCCGGTC CCGGGGCGGT GAAGACGGCT 500

FIG. 31A

42/164

plkPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890
TCCGGAAGT TCTGGAAGC GCGGCTGGC CTCCTGGCA GCAGGGCTG GGGGCTGGC CTGATCTCA GGACGGCC GTCGACGGG GGCATCGAC									
600									
GGGACGACGA GGGCGGAGC CGCTCGTGA CCGCGACTC GTACCAGGG GCCCAGGAG TCGTGAGAG CGATCCGGC CTGCGACCG CGCGTTTGG									
700									
GGCGACCTCC CACTGCCTA TCAGGGCCG CGACTGCCAG TTCGTGATT CCAGGAGACT CCGGAGATC TGGTCTGTA TCGCGAAGG AACGTTTCCG									
800									
ACGACGGTGT CGATATCGC CGGATGGG AAGTCGAGGA AATCACCCTG GAATACGGTG ACCCTCTCC CTTCGAATTT CCGCGGCACA TCGCGGGCC									
900									
AGTCGGGTC CATCTCCAC ACCGTCACG TGTCGAAGGA GGCACCAAC TCTCTGGTTA TCGCGCCCTT TCGGGGGCG ATTTCGAGAA CGTTCCTAGC									
1000									

ApoI
BstBI
▼ ▼

EcoRV
▼
BsmI
▼

FIG. 31B

43/164

plkPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890

GTCCCCCTCG ACATGCGTGA CGAGATWGG CACGGCTCTG TCGTCTGTGA GGAAGTCTTG GCCTAATTCG CCGCGAAGGG TGTCCGGTTC CGCTCGACTC 1100

GCTATGGAGT CGAGCATTCG CATgaacgat cccctccctg gatgccgtgg tcaatggact tggcacggac catacctcac ggtccgtcgg acgaccggag 1200

XmnI



aagaagtcca cgcacgggcg ttccggagta cgggagttgt gaacggccgc gacgaagtcg gtcgcggtc ggcggggcgt gacgagcgag gtccggagga 1300

acgcgacgaa gcagccgaac cccaagtga gtgcgacgga gtgacattgg gggcatacgg agggttgtcg tacggagcgc actcaacgag gctccaggag 1400

ggaggggttg aaccgcgcgc cgactggcct tcgccgcccg cgcggccgga gtagtgcata tcgggggtga aatcaagcca ttccccccggg atcggctgtt 1500

FIG. 31C

44/164

plPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890

accatccct ttacctggcg tggatttccc aaccttgggt atagagcggg agacgacgcg acaccatgga gaccacgcac accacgagcg ccaccccccg 1600

gccatccga caaggggggt ccggctcgcc tcccgacacc catggcctgg ggtacacgcc aggtataggg ggaacgtagg gggagcatag ggggggtgcc 1700

ctggggttgg gtgaaagcgc ggcttccgga gacgggagcgg gatgtcttca gccggaatta ccaggaccgg tgcgagaaca ccggtgacag ggcgtggggc 1800

ggcagcgtgg gacacggggg aagtgcgggt ccgacggggg ttgccccctg ccggccccga tcatgaggag cactccttct ctggtgctcc tacgggtgat 1900

EcoRI
 ApoI

XmnI

gtgcggcggc aattgattcg tggagagatg tgcacagtgt ccaagagtga gtccgaggaa ttccgtgtccg tgcgagaaca cgccgggtcc gcgcacggca 2000

FIG. 31D

45/164

pikPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890

CAGCGGAACC CGTCGCCGTC GTCGGCATCT CCTGCCGGGT GCCCGGGCC CGGACCCGA GAGAGTTCTG GGAATCCTG GCGGCAGGCG GCCAGGCCGT 2100
A E P V A V V G I S C R V P G A R D P R E F W E L L A A G G Q A V

CACCGACGTC CCCCGGACC GCTGGAGCG CGGGACTTC TAGACCCGG ACGCTCCGC CCCGGCCGC TCGAACAGCC GGTGGGCGG GTTCATCGAG 2200
T D V P A D R W N A G D F Y D P D R S A P G R S N S R W G G F I E

GACGTCGACC GGTTCGACG CGCCTTCTTC GGCATCTCG CGCGGAGATG GACCCGACG AGCGGCTCGC CCTGGAGCTG GGCTGGGAGG 2300
D V D R F D A A F F G I S P R E A A E M D P Q Q R L A L E L G W E A

CCCTGGAGCG CGCGGGATC GACCGTCCT CGCTCACCG CACCCGCACC GCGTCTTTC GCGGCGCAT CTGGGACGAC TACGCCACCC TGAAGCACCG 2400
L E R A G I D P S S L T G T R T G V F A G A I W D D Y A T L K H R

CCAGGGGCGC GCGCGGATCA CCCCGCACAC CGTCACCGG CTCACCGCG GCATCATCGC GAACCGACTC TCGTACACG TCGGGCTCCG CGGCCCCGAGC 2500
Q G G A A I T P H T V T G L H R G I I A N R L S Y T L G L R G P S

FIG. 31E

46/164

plkPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890

ATGTCGTGTCG ACTCCGGCCA GTCCTCGTCG CTCGTGCGCG TCCACCTCGC GTGCGAGAGC CTGCGGGCGG GCGAGTCCGA GCTCGCCCTC GCCGGCGCGG 2600
 M V V D S G Q S S S L V A V H L A C E S L R R G E S E L A L A G G V

TCTCGCTCAA CCTGGTGGCG GACAGCATCA TCGGGGGGAG CAAGTTCGGC GGCCTCTCCC CCGACGGCGG CCGCTACACC TTCGACGGC GCGCCAACGG 2700
 S L N L V P D S I I G A S K F G G L S P D G R A Y T F D A R A N G

SnaBI



CTACGTACGC GCGAGGGCG GCGGTTCGT CGTCTGAAG CGCCTCTCCC GCGCGTCCG CCGAGGGGAC CCGTGTCTCG CCGTGTCTCG GGGCAGCGCC 2800
 Y V R G E G G G F V V L K R L S R A V A D G D P V L A V I R G S A

GTCAACACG GCGGCGCGC CCAGGGCATG ACGACCCCG ACGGCGAGG GTGCTCCGG AGGCCACGA GCGGGCGGG ACCGCGCGG 2900
 V N N G G A A Q G M T T P D A Q A Q E A V L R E A H E R A G T A P A

CCGACGTGCG GTACGTGAG CTGCACGGCA CCGGCACCCC CCGTGGCGAC CCGATCGAGG CCGTGTGCT CCGGCGCGCC CTCGGCACCG GCCGCCCGGC 3000
 D V R Y V E L H G T G T P V G D P I E A A A L G A A L G T G R P A

FIG. 31F

47/164

plkPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890
CGGACAGCCG	CTCCTGGTGG	GCTCGGTCAA	GACGAACATC	GGCCACCTGG	AGGGCGCGGC	CGGCATCGCC	GGCCTCATCA	AGGCCGTCTT	GGCGGTCCGC
G Q P	L L V G	S V K T	N I G H	L E G A	A G I A	G L I K	A V L A	V R	
3100									
GGTCGGCGGC	TGCGCGCCAG	CCTGAACTAC	GAGACCCCGA	ACCGGCGGAT	CCCGTTCGAG	GAACCTGAACC	TCCGGGTGAA	CACGGAGTAC	CTGCCGTGGG
G R A L	P A S L	N Y E T	P N P A	I P F E	E L N L	R V N T	E Y L P	W E	
3200									
AGCCGGAGCA	CGACGGGCAG	CGGATGGTGG	TCGGCGTGTC	CTCGTTCGGC	ATGGCGGGCA	CGAACCGGCA	TGTCGTGCTC	GAAGAGGCC	CCGGGGTTG
P E H	D G Q R	M V V G	V S S F	G M G G	T N A H	V V L E	E A P G	G C	
3300									
TCGAGGTGCT	TCGTCGTGG	AGTCGACGGT	CGGCGGGTGG	GGCGTGGCG	GCGGTGTGGT	GCCGTGGGTG	GTGTCGGCGA	AGTCCGCTGC	CGCGCTGGAC
R G A	S V V E	S T V G	G S A V	G G G V	V P W V	V S A K	S A A L	D	
3400									
GCGCAGATCG	AGCGGCTTGC	CGCGTTCGCC	TCGCGGGATC	GTACGGATGG	TGTGACGCG	GGCGTGTGCG	ATGCGGGTGC	TGTGATGCG	GGTGTGTGCG
A Q I E	R L A A	F A S R	D R T D	G V D A	G A V D	A G A V	D A G A	V A	
3500									

FIG. 31G

48/164

plPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890
CTCGCGTACT	GGCGGGCGGG	CGTGCTCAGT	TCGAGCACCG	GGCGGTCGTC	GTCGGCAGCG	GGCGGACGCA	TCTGGCGGCA	GGCGTGCGG	CGCCTGAGGG
R V L	A G G	R A Q	F E H	R A V	V V G	S G P	D D L	A A A	L A A
3600									
TCTGGTCCGG	GGCGTGGCTT	CCGGTGTCGG	GCGAGTGCGG	TTGGTGTTCC	CCGGGCAGGG	CACGCACTGG	GCCGGCATGG	GTGCCGAACT	GCTGGACTCT
L V R	G V A	S G V	G R V	A F V	F P G	Q G T	Q W A	G M G	A E L
3700									
TCGCGGTGT	TCGCGGCGGC	CATGGCCGAA	TGCGAGGCGG	CACCTCTCCC	GTCAGTCGAC	TGGTCGCTGG	AGGCCGTCGT	ACGGCAGGCC	CCCGGTGCGC
S A V	F A A	A M A	E C E	A A L	S P Y	V D W	S L E	A V V	R Q A
3800									
CCACGCTGGA	GGGGTCCGAT	GTCGTGCAGC	CTGTGACGTT	CGCGTTCATG	GTCCTGCTGG	CTCGCGTGTG	GCAGCACCCAC	GGGGTGACGC	CCCAGGCGGT
T L E	R V D	V V Q	P V T	F A V	M V S	L A R	V W Q	H H G	V T P
3900									
CGTCGGCCAC	TCGCAGGGCG	AGATCGCCGC	CGCGTACGTC	GGCGGTGCCC	TGAGCCTGGA	CGACGCGGCT	CGTGTCTGTA	CCCTGCGCAG	CAAGTCCATC
V G H	S Q G	E I A	A A Y	V A G	A L S	L D D	A A R	V V T	L R S
4000									

FIG. 31H

49/164

pikPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890
GGCGCCACC TCGCCGGCAA GCGCGGCATG CTGTCCCTCG CGCTGAGCGA GGACGCCGTC CTGGAGCGAC TGGCCGGGTT CGACGGGCTG TCCGTGCGCG 4100									
A A H L A G K G G M L S L A L S E D A V L E R L A G F D G L S V A A									
CTGTGAACGG GCCACACGCC ACGTGGTCT CCGGTGACCC CGTACAGATC GAAGAGCTTG CTCGGGCGTG TGAGGCCGAT GGGGTCCGTG CGCGGGTCAT 4200									
V N G P T A T V V S G D P V Q I E E L A R A C E A D G V R A R V I									
TCCCCTCGAC TACCGGTCCC ACAGCCGGCA GGTGAGATC ATCGAGAGCG AGCTCGCGA GGTCTCGCC GGGCTCAGCC CGCAGGTCC CGCGGTGCGG 4300									
P V D Y A S H S R Q V E I I E S E L A E V L A G L S P Q A P R V P									
TTCTTCTCGA CACTCGAAGG CGCCTGGATC ACCGAGCCCG TGCTCGACGG CGGTACTGG TACCGCAACC TCGGCCATCG TGTGGGCTTC GCCCGGCGCG 4400									
F F S T L E G A W I T E P V L D G G Y W Y R N L R H R V G F A P A V									
TCGAGACCCT GGCCACCGAC GAGGCTTCA CCGACTTCGT CGAGGTGAGC GCCACCCCG TCCTCACCAT GGCCCTCCCC GGGACCGTCA CCGGTCTGGC 4500									
E T L A T D E G F T H F V E V S A H P V L T M A L P G T V T G L A									

FIG. 311

50/164

plkPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890
MaeI BfaI ▼									
GACCTGCGT	CGCGACAACG	GCGGTACAGG	CGGCCTAGTC	GCCTCCCTCG	CCGAAGCATG	GGCCAACGGA	CTCGCGGTCTG	ACTGAGAGCCC	GCTCCTCCTCCC
T L R	R D N G	G Q D	R L V	A S L A	E A W	A N G	L A V D	W S P	L L P
MluI ▼									
TCCGCGACCG	GCCACCACTC	CGACCTCCTCCC	ACCTACGCGT	TCCAGACCGA	GCGCCACTGG	CTGGGCGAGA	TCGAGGCGCT	CGCCCCGCGG	GGCGAGCCGG
S A T G	H H S	D L P	T Y A F	Q T E	R H W	L G E I	E A L	A P A	G E P A
MluI ▼									
CGGTGAGCC	CGCCGTCTC	CGCACGGAGG	CGGCGAGCTC	GACCGGACG	AGCAGCTGCG	CGTGATCCTG	GACAAGGTCC	GGGCGCAGAC	4800
V Q P	A V L	R T E A	A E P	A E L	D R D E	Q L R	V I L	D K V	R A Q T
MluI ▼									
GGCCCAGGTG	CTGGGGTACG	CGACAGGCGG	GCAGATCGAG	GTCGACCGGA	CCTTCCGTGA	GGCGGTTGTC	ACCTCCCTGA	CCGGCGTGGA	CCTGCGCAAC
A Q V	L G Y A	T G G	Q I E	V D R	T F R E	A G C	T S L	T G V D	L R N
MluI ▼									
CGGATCAACG	CCGCCTTCGG	CGTACGGATG	GCGCCGTCCA	TGATCTTCGA	CTTCCCACAC	CCGAGGCTC	TCGCGGAGCA	GCTGCTCCTC	GTCGTGCACG
R I N A	A F G	V R M	A P S M	I F D	F P T	P E A L	A E Q	L L L	V V H G
MluI ▼									
ApaLI ▼									
MluI ▼									

FIG. 31J

51/164

plkPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890

GGGAGGCGGC GCGGAACCG GCGGTGGCG AGCCGGCTCC GGTGGCGCG GCGGTGCGG TCGACGAGCC GGTGGCGATC GTCGGCATGG CCTGCCGCCT 5100
E A A A N P A G A E P A P V A A A G A V D E P V A I V G M A C R L

GCCCGGTGGG GTCGCTCGC CGGAGGACCT GTGGCGGCTG GTGGCCGGCG GCGGGGACGC GATCTCGGAG TTCCCGCAGG ACCGCGGCTG GGACGTGGAG 5200
P G G V A S P E D L W R L V A G G G D A I S E F P Q D R G W D V E

BamHI
▼

GGGCTGTACC ACCCGATCC GGAGCACCCC GGCACGTGCT ACGTCGCCCA GGGCGGTTTC ATCGAGAACG TCGCCGGCTT CGACGCGGCC TTCTTCGGGA 5300
G L Y H P D P E H P G T S Y V R Q G G F I E N V A G F D A A F F G I

TCTCGCCTCG CGAGGCGCTC GCCATGGACC CGCAGCAGCG GCTCCTCCTC GAAACCTCCT GGGAGGCCGT CGAGGACGCC GGGATCGACC CGACCTCCCT 5400
S P R E A L A M D P Q Q R L L L E T S W E A V E D A G I D P T S L

GCGGGGACGG CAGGTCGGCG TCTTCACTGG GCGATGACC CACGAGTACG GGCCGAGCCT GCGGGACGCC GGGGAAGGCC TCGACGGCTA CCTGCTGACC 5500
R G R Q V G V F T G A M T H E Y G P S L R D G G E G L D G Y L L T

FIG. 31K

52/164

plkPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890

GGCAACACGG CCAGCGTGAT GTCGGGCGC GTCTCGTACA CACTGGGCCT TGAGGGCCC GCCCTGACGG TGGACACGGC CTGCTCGTCG TCGTGTGTCG 5600
 G N T A S V M S G R V S Y T L G L E G P A L T V D T A C S S S L V A

CCCTGCACCT CGCCGTGCAG GCCCTGGCA AGGGCGAGGT CGACATGGCG CTCGGCGCG GCGTGGCCGT GATGCCACG CCGGGATGT TCGTCGAGTT 5700
 L H L A V Q A L R K G E V D M A L A G G V A V M P T P G M F V E F

XmnI



CAGCCGGCAG CGCGGGCTGG CCGGCGACGG CCGGTCGAAG GCGTTCGCCG CGTCGGCGGA CGGCACCAGC TGGTCCGAGG GCGTCGGCGT CCTCCTCGTC 5800
 S R Q R G L A G D G R S K A F A A S A D G T S W S E G V G V L L V

GAGCCCTGT CGGACGCCC CGGCAACGGA CACCAGGTCC TCGCGGTGCT CGCGGCGCAG GCCTTGAACC AGGACGGCGC GAGCAACGGC CTCACGGCTC 5900
 E R L S D A R R N G H Q V L A V V R G S A L N Q D G A S N G L T A P

CGAACGGGCC CTCGACGAG CGCGTCATCC GCGCGCGCT GCGGACGCC CGGCTGACGA CCTCCGACGT GAGGTCGTC GAGGCACAG GCACGGGCAC 6000
 N G P S Q Q R V I R R A L A D A R L T T S D V D V V E A H G T G T

FIG. 31L

53/164

plkPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890
GGGACTCGGC	GACCCGATCG	AGGCGCAGGC	CCTGATCGCC	ACCTACGGCC	AGGCGCGTGA	CGACGAACAG	CCGCTGCGCC	TCGGGTGCGT	GAGTCCAAAC
R L G	D P I E	A Q A	L I A	T Y G Q	G R D	D E Q	P L R L	G S L	K S N
6100									
ATCGGGCACA	CCGAGGCCGC	GGCCGGCGTC	TCCGGGTCTA	TCAAGATGGT	CCAGGCGGATG	CGCCACGGAC	TGCTGCCGAA	GACGCTGCAC	GTCGACGAGC
I G H T	Q A A	A G V	S G V I	K M V	Q A M	R H G L	L P K	T L H	V D E P
6200									
CCTCGGACCA	GATCGACTGG	TCGGCTGGCG	CCGTGGAACT	CCTCACCAG	GCCGTCGACT	GGCCGGAGAA	GCAGGACGGC	GGGCTGCGCC	GGGCGCGCGT
S D Q	I D W	S A G A	V E L	L T E	A V D W	P E K	Q D G	G L R R	A A V
6300									
CTCTCTCTTC	GGGATCAGCG	GCACCAATGC	GCATGTGTG	CTCGAAGAGG	CCCGGTGGT	TGTCGAGGCT	GCTTCGGTCTG	TCGAGCCGTC	GGTTGGCGGG
S S F	G I S G	T N A	H V V	L E E A	P V V	V E G	A S V V	E P S	V G G
6400									
TCGGCGGTTCG	GCGGCGGTGT	GACGCTTGG	GTGGTGTCTG	CGAAGTCCGC	TGCCGCGCTC	GACGCGCAGA	TCGAGCGGCT	TGCCGATTC	GCCTCGCGGG
S A V G	G G V	T P W	V V S A	K S A	A A L	D A Q I	E R L	A A F	A S R D
6500									

BamI



FIG. 31M

54/164

pkPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890
ATCGTACGGA	TGACGCCGAC	GCCGGTCTG	TCGACGCGG	CGCTGTCGCT	CACGTACTGG	CTGACGGGG	TGCTCAGTTC	GAGCACCGGG	CCGTGCGGCT
R T D D A D	A G A V D A G	A V A H V L A	D G R A Q F	E H R A V A	L				
6600									
GGCGCCGGG	GCGACGACC	TCGTACAGGC	GCTGGCCGAT	CCGACGGGC	TGATACGGG	AACGGCTTCC	GGTGTGGGC	GAGTGGCGTT	CGTGTTCCTC
G A G A D D L	V Q A L A D	P D G L I R G	T A S G V G R	V A F V F P					
6700									
3GTCAGGGCA	CGCAGTGGC	TGGCATGGT	GCCGAACCTGC	TGACTCTTC	CGCGGTGTC	GCGGGGCCA	TGGCCGAGTG	TGAGGCCCGG	CTGTCCCGCT
3 Q G T Q W A	G M G A E L L	D S S A V F A	A A M A E C E	A A L S P Y					
6800									
ACGTCGACTG	GTGCTGGAG	GCCGTCGTAC	GCCAGGCCCC	CGGTGCGCC	ACGCTGGAGC	GGTCTGATGT	CGTGCAGCCT	GTGACGTTTG	CCGTGATGGT
V D W S L E	A V V R Q A P	Q A P G A P	T L E R V D V	V Q P V T F A	V M V				
6900									
CTCGCTGGCT	CGCGTGTGG	AGCACCACGG	TGTGACGCC	CAGGCGGTG	TCGGCCACTC	GCAGGGCGAG	ATCGCGCCG	CGTACGTCGC	CGGAGCCCTG
S L A R V W Q	H H G V T P	Q A V V G H S	Q G E I A A A	Y V A G A L					
7000									

XmnI
▼

FIG. 31N

55/164

plkPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890

SphI
▼

CCCCCTGGACG ACGCCGCCCG CGTCGTACAC CTGCGCAGCA AGTCCATCGC CCCCCACCTC GCGGCAAGG GCGGCATGCT GTCCCTCGCG CTGACGAGG 7100
P L D D A A R V V T L R S K S I A A H L A G K G G M L S L A L N E D

ACGCCGTCTT GGAGCGACTG AGTGACTTCG ACGGGCTGTC CGTCGCCGCC GTCAACGGGC CCACGCCAC TGTGCTGTCG GGTGACCCCG TACAGATCGA 7200
A V L E R L S D F D G L S V A A V N G P T A T V V S G D P V Q I E

MluI
▼

AGAGCTTGCT CAGGCGTGCA AGGCGGACGG ATTCCGCGCG CGGATCATTG CCGTCGACTA CGGTCGCCAC AGCCGGCAGG TCGAGATCAT CGAGAGCGAG 7300
E L A Q A C K A D G F R A R I I P V D Y A S H S R Q V E I I E S E

CTCGCCGAGG TCCTGCGCGG TCTCAGCCCG CAGGCCCGCG GCGTCGCGTT CTTCCTGACG CTCGAAGGCA CCTGGATCAC CGAGCCCGTC CTCGACGGCA 7400
L A Q V L A G L S P Q A P R V P F F S T L E G T W I T E P V L D G T

KpnI
▼
Acc65I
▼

CCTACTGGTA CCGCAACCTC CGTCACCGCG TCGGCTTCGC CCGCGCCATC GAGACCCCTGG CCGTCGACGA GGGCTTCACG CACTTCGTCG AGGTCAGCGC 7500
Y W Y R N L R H R V G F A P A I E T L A V D E G F T H F V E V S A

FIG. 310

56/164

pkPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890

CCACCCCGTC CTCACCATGA CCTTCCCGCA GGCCTCGGCA CCTCCGTCG CGAACAGGGA GGCCAGAGC GTCTGGTCAC CTGCTCGCC 7600
H P V L T M T L P E T V T G L G T L R R E Q G G Q E R L V T S L A

Eco47III



GAGCGTGGG TCAACGGGT TCCGTGGCA TGGACTTCG TCTGCCCGC CACGGCCTCC CGCCCGGTC TGCCACCTA CGCCTTCCAG GCCGAGCGT 7700
E A W V N G L P V A W T S L L P A T A S R P G L P T Y A F Q A E R Y

XhoI

PaeR7I

MscI

BalI



ACTGGCTCGA GAACACTCCC GCCGCCCTGG CCACCGGCGA CGACTGGGC TACCGCATCG ACTGGAAGCG CCTCCCGGCC GCCGAGGGGT CCGAGCGCAC 7800
W L E N T P A A L A T G D D W R Y R I D W X R L P A A E G S E R T

CGGCTGTCC GGCCGCTGGC TCGCCGTCAC GCCGAGGAC CACTCCGCGC AGGCGCGCGC CGTGTCTACC GCGCTGGTCG ACGCCGGGC GAAGGTGAG 7900
G L S G R W L A V T P E D H S A Q A A A V L T A L V D A G A K V E

GTGCTGACG CGCGGGCGGA CGACGACCGT GAGGCCCTCG CGGCCGGGT CACCGCACTG ACGACGGTG ACGGCTTCA CGGCGTGGT TCGCTCCTCG 8000
V L T A G A D D D R E A L A A R L T A L T T G D G F T G V V S L L D

FIG. 31P

57/164

pkPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890

ACGGACTCGT ACCGACAGGTC GCCTGGGTCC AGGCGCTCGG CGACGCCCGGA ATCAAGGCGC CCTGTGGTC CGTACCCAG GCGCGGTCT CCGTCGGACG 8100
 G L V P Q V A W V Q A L G D A G I K A P L W S V T Q G A V S V G R

TCTCGACACC CCGGCCGACC CCGACCGGC CATGCTCTGG GGCCTCGGC GGTGTCGCG CCTGAGCAC CCGAACGCT GGGCCGGCCT CGTCGACCTC 8200
 L D T P A D P D R A M L W G L G R V V A L E H P E R W A G L V D L

BsaBI



CCGCCCCAGC CCGATGCGC CGCCTCGCC CACTCTCTCA CCGACTCTC CGCGCCACC GCGAGGACC AGATCGCCAT CCGCACCACC GGACTCCACG 8300
 P A Q P D A A A L A H L V T A L S G A T G E D Q I A I R T T G L H A

CCGCGCGCCT CGCCCGGCA CCCCTCCACG GACGTGCGC CACCCGGAC TGGCAGCCCC ACGGACCGT CCTCATCACC GCGCGCACCG GAGCCCTCGG 8400
 R R L A R A P L H G R R P T R D W Q P H G T V L I T G G T G A L G

CAGCCACGCC GCACGCTGGA TGGCCACCA CCGAGCGGA CACCTCCTCC TGTGAGCG CAGCGCGGA CAAGCCCCCG GAGCCACCA ACTCACCGCC 8500
 S H A A R W M A H H G A E H L L L L V S R S G E Q A P G A T Q L T A

FIG. 31Q

58/164

plkPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890

GAACTCACCG CATGGGGGC CCGCGTCACC ATCGCCGCCT GCGAGTCGC CGACCCCCAC GCCATGGCA CCTCTCTGA CGCCATCCC GCCGAGACG 8600
E L T A S G A R V T I A A C D V A D P H A M R T L L D A I P A E T P

CCCTCACCGC CGTGTCCAC ACCGCGGCG CGCTCGACGA CGGCATCGT GACACGCTGA CCGCGAGCA GGTCCGCGG GCCACCGTG CGAAGGCCGT 8700
L T A V V H T A G A L D D G I V D T L T A E Q V R R A H R A K A V

MluI
▼

CGGCGCCTCG GTGCTCGAC AGCTGACCG GGACCTCGAC CTGACGCGT TCGTCTCTT CTCGTCCGTG TCGAGCACTC TGGGCATCCC CGGTCAGGCG 8800
G A S V L D E L T R D L D L D A F V L F S S V S S T L G I P G Q G

AACTACGCCC CGCACACGC CTACCTCGAC GCCCTGCGG CTGCGCGCG GGCACCGGC CGGTCCGCG TCTCGGTGC CTGGGACCG TGGACGCTG 8900
N Y A P H N A Y L D A L A A R R R A T G R S A V S V A W G P W D G G

GCGGCATGGC CGCCGCTGAC GCGGTGGCG AGCGGCTGCG CAACACGCG GTGCCCGCA TGGACCCGA ACTCGCCCTG GCCGCACTGG AGTCCGCGCT 9000
G M A A G D G V A E R L R N H G V P G M D P E L A L A A L E S A L

FIG. 31R

59/164

plkPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890

CGGCCGGGAC GAGACCGCGA TCACCGTCGC GGACATCGAC TGGGACCGCT TCTACCTCGC GTACTCCTCC GGTGCGCCGC AGCCCCCTCGT CGAGGAGCTG 9100
G R D E T A I T V A D I D W D R F Y L A Y S S G R P Q P L V E E L

BstXI



CCCGAGGTGC GCGCGATCAT CGACGCACGG GACAGCGCCA CGTCCGGACA GGGCGGGAGC TCCGCCCAGG GCGCCAACCC CCTGGCCGAG CGGCTGGCCG 9200
P E V R R I I D A R D S A T S G Q G G S S A Q G A N P L A E R L A A

CCGCGGCTCC CCGCGAGCGT ACGGAGATCC TCCTCGGTCT CGTACGGGCG CAGGCGCGCG CCGTGCTCCG GATGCTTCG CCGGAGGACG TCGCCGCCGA 9300
A A P G E R T E I L L G L V R A Q A A A V L R M R S P E D V A A D

CCGCGCCTTC AAGGACATCG GCTTCGACTC GCTCGCCGGT GTCGAGCTGC GCAACAGGCT GACCCGGCG ACCGGGCTCC AGCTGCCCGC GACGCTCGTC 9400
R A F K D I G F D S L A G V E L R N R L T R A T G L Q L P A T L V

TTGACCAACC CGACGCGCGT GGCCTCTGTG TCGCTGTCTC GCAGCGAGTT CCTCGGTGAC GAGGAGACGG CCGACGCCCG GCGGTCCGCG GCGCTGCCCG 9500
F D H P T P L A L V S L L R S E F L G D E E T A D A R R S A A L P A

FIG. 31S

60/164

plkPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890

BsaBI

CGACTGTGCG TGCCGGTGCC GGCGCCGGCG CCGGCACCGA TGCCGACGAC GATCCGATCG CGATCGTCCG GATGAGCTGC CGTACCCCG GTGACATCCG 9600
T V G A G A G A G A G T D A D D A D D D P I A I V A M S C R Y P G D I R

CAGCCCGGAG GACTGTGGC GGATGCTGTC CGAGGGCGGC GAGGGCATCA CGCGTTCCC CACCGACCGC GGCTGGGACC TCGACGGCCT GTACGACGCC 9700
S P E D L W R M L S E G G E G I T P F P T D R G W D L D G L Y D A

GACCCGGACG CGCTCGACG GCGGTACGTC CCGAGGGCG GGTTCCTGCA CGACGCGGCC GAGTTCGACG CGGAGTTCTT CGGCGTCTCG CCGCGCGAGG 9800
D P D A L G R A Y V R E G G F L H D A A E F D A E F F G V S P R E A

MscI
Bali

CGCTGGCCAT GGACCCGACG CAGCGGATGC TCCTGACGAC GTCTGGGAG GCCTTCGAGC GGGCCGGCAT CGAGCCGGCA TCGCTGCGCG GCAGCAGCAC 9900
L A M D P Q Q R M L L T T S W E A F E R A G I E P A S L R G S S T

CGGTGTCTTC ATCGGCTCTT CCTACCAGGA CTACGCGGCC CCGTCCCGA ACGCCCCCG TGGCGTGGAG GGTACCTGC TGACCGGCAG CACGCCGAGC 10000
G V F I G L S Y Q D Y A A R V P N A P R G V E G Y L L T G S T P S

FIG. 31T

61/164

plkPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890

GTCGCGTCGG GCCGTATCGC GTACACCTTC GGTCTCGAAG GGCCCGCGAC GACCGTCGAC ACCGCTGCT CGTCGTCGCT GACCGCCCTG CACCTGGCGG 10100
 V A S G R I A Y T F G L E G P A T T V D T A C S S S L T A L H L A V

▼
 ApaLI

TGCGGGCGCT GCGCAGCGGC GAGTGCACGA TGGCGCTCGC CGGTGGCGTG GCGATGATGG CGACCCCGCA CATGTTCTGT GAGTTCAGCC GTCAGCGGCG 10200
 R A L R S G E C T M A L A G G V A M M A T P H M F V E F S R Q R A

GCTCGCCCCG GACGCGCGCA GCAAGGCCTT CTCGGCGGAC GCCGACGGT TCGCGCGCGC GGAGGCGTC GGCCTGCTGC TCGTGGAGCG GCTCTCGGAC 10300
 L A P D G R S K A F S A D A D G F G A A E G V G L L L V E R L S D

▼
 KpnI
 Acc65I

GCGCGGCGCA ACGGTACCC GGTGCTCGCC GTGTCCGCG GTACCGCGT CAACCAGGAC GGCGCCAGCA ACGGGCTGAC CGCGCCCAAC GGACCCCTCGC 10400
 A R R N G H P V L A V V R G T A V N Q D G A S N G L T A P N G P S Q

AGCAGCGGGT GATCCGGCAG GCGCTCGCG AGCCCGGCT GGACCCCGGC GACATCGACG CCGTCGAGAC GCACGGCAGC GGAACCTGCG TGGCGGACCC 10500
 Q R V I R Q A L A D A R L A P G D I D A V E T H G T G T S L G D P

FIG. 31U

62/164

plPKS Sequence

	20	30	40	50	60	70	80	90	100
1	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890
2	CAGGGCCTCC	AGGCCACGTA	CGGCAAGGAG	CGGCCCGCGG	AACGGCCGCT	CGCCATCGGC	TCCGTGAAGT	CCAACATCGG	ACACACCCAG
3	Q G L Q	A T Y G	K E R P	A E R P	L A I G	S V K S	N I G	H T Q	
4	GTGCGGCGGG	CATCATCAAG	ATGGTCCTCG	CGATGCGCCA	CGGCACCCCTG	CCGAAGACCC	TCCACGCCGA	CGAGCCGAGC	CCGCACGTCG
5	A A G I	I K M V	L A M R	H G T L	P K T L	H A D	E P S	P H V D	
6	CAGCGGCCTG	GCCCTCGTCA	CCGAGCCGAT	CGACTGGCCG	GCCGGCACCG	GTCCGGCGCG	CGCGCGCGTC	TCCTCCTTCG	GCATCAGCGG
7	S G L A	L V T E	P I D W	P A G T	G P R R	A A V	S S F G	I S G	
8									
9									
10									
11									
12									
13									
14									
15									
16									
17									
18									
19									
20									
21									
22									
23									
24									
25									
26									
27									
28									
29									
30									
31									
32									
33									
34									
35									
36									
37									
38									
39									
40									
41									
42									
43									
44									
45									
46									
47									
48									
49									
50									
51									
52									
53									
54									
55									
56									
57									
58									
59									
60									
61									
62									
63									
64									
65									
66									
67									
68									
69									
70									
71									
72									
73									
74									
75									
76									
77									
78									
79									
80									
81									
82									
83									
84									
85									
86									
87									
88									
89									
90									
91									
92									
93									
94									
95									
96									
97									
98									
99									
100									

Bsu36I



FIG. 31V

63/164

plPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890

CCGACGTGGC CCTGCCGGCC GAGGTGCGCG GTGCCGAGGG CGCGCGGTTC GGCCTTCACC CGCGCTGCT CGACGCCGCC GTGCAGGCGG CCGGTGCGGG 23600
D V A L P A E V A G A E G A R F G L H P A L L D A A V Q A A G A G

CGGGGGGTT CGGCGGGGC ACGGGCTGC CGTTGCGCTG GAGCGGATC TCCTGTACG GGTGGGCGC ACCGCCCTCC GCGTGGGCT GGCCCCCGCC 23700
R G V R R G H A A A V R L E R D L L Y A V G A T A L R V R L A P A

GGCCCGGACA CGGTGTCCGT GAGCGCGCC GACTCCTCG GGCAGCGGT GTTCGCGCG GACTCCCTCA CGGTGTGTC CGTCGACCCC GCGCAGCTGG 23800
G P D T V S V S A A D S S G Q P V F A A D S L T V L P V D P A Q L A

CGGCCTTCAG CGACCCGACT CTGGACGCG TGCACCTGCT GGAGTGACC GCCTGGAGC GTGCCGCGCA GGCCCTGCCC GCGCGGGTGG TGCTGGGCGG 23900
A F S D P T L D A L H L L E W T A W D G A A Q A L P G A V V L G G

CGACGCCGAC GGTCTGCGCG CGCGGTGGC ACCAGGTCC TGTCCTTCCC GGACCTTACG GACCTGTGG AGGCCGTGA CCGGGGCGAG 24000
D A D G L A A A L R A G G T E V L S F P D L T D L V E A V D R G E

FIG. 31W

64/164

plkPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890

SphI
▼

GCCTCTTCC GGCTCGTCCA GAGCTGGGCG ATGCGGCGCG CGCACTGCT CGGTCACTCG GTCGGCGAGA TCGCGCGCG GCACGTGCGC GGTGTGTCT 11600
A L F R L V E S W G M R P A A L L G H S V G E I A A A H V A G V F S

CGCTGCGCGA CGCGGCGCGC CTGCTGCGCG CGCGGCGCG GCTCATGCA GAGTCTCCCG CCGGTGGCGC GATGCTGCGC GTCCAGGCGC CGAGAGGACGA 11700
L A D A A R L V A A R G R L M Q E L P A G G A M L A V Q A A E D E

GATCCGCGTG TGGCTGAGA CGGAGGAGCG GTACGCGGGA CGTCTGGAC TCGCCCGCGT CAACGGCCCC GAGGCGCGC TCCTGTCCCG CGACCGGAC 11800
I R V W L E T E E R Y A G R L D V A A V N G P E A A V L S G D A D

SphI
▼

GCGGCGCGG AGGCGGAGC GTACTGGTCC GGGCTCGCC GCAGGACCG CGGCTGCGG GTCAGCCAG CCTTCCACTC CGGCACATG GACGGCATGC 11900
A A R E A E A Y W S G L G R R T R A L R V S H A F H S A H M D G M L

TGACCGGGT CGCGGCGCTC CTGAGAGCG TGGAGTTCCG GCGCCCTCC CTGACGTGG TCTGAACTG CACCGGCGTG GCGCGCGCG CGAGGACCT 12000
D G F R A V L E T V E F R R P S L T V V S N V T G L A A G P D D L

FIG. 31X

65/164

plkPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890

ScaI



GTGCGACCCC GAGTACTGGG TCCGGCAGCT CCGCGGCACC GTCCGCTTCC TCGACGGCGT CCGTGTCCTG CCGGACCTCG GCGTCCGGAC CTGCTGGAG 12100
 C D P E Y W V R H V R G T V R F L D G V R V L R D L G V R T C L E

CTGGGCCCCG ACGGGTCTT CACCGCCATG GCGGCGGACG GCCTCGCGGA CACCCCGCG GATTCGGCTG CCGGCTCCCC CGTGGCTCT CCGCCCGGCT 12200
 L G P D G V L T A M A A D G L A D T P A D S A A G S P V G S P A G S

CTCCCGCCGA CTCGCGCGCC GCGCGGCTCC GGCCCGGCCC GCTGCTCTG GCGCTGCTG CCGGCAAGCG GTCGGAGACC GAGACCGTCG CGGACGCCCT 12300
 P A D S A A G A L R P R P L L V A L L R R K R S E T E T V A D A L

CGGCAGGGCG CAGGCCACG GCACCGGACC CGACTGGCAC GCCTGGTTCG CCGGCTCCGG GCGGCACCGC GTGGACCTGC CCACGTACTC CTTCCGGCGC 12400
 G R A H A H G T G P D W H A W F A G S G A H R V D L P T Y S F R R

GACCGTACT GGCTGGACG CCCGGCGGCC GACACCGCGG TGGACACCGC CGGCCTCGGT CTCGGCACCG CCGACCAACC GCTGCTCGGC GCGTGGTCA 12500
 D R Y W L D A P A A D T A V D T A G L G L G T A D H P L L G A V V S

FIG. 31Y

66/164

plkPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890

GCCTTCCGGA CCGGACGCGC CTGTGCTCA CCGGCGCCT CTCCCTGCGC ACCACCCGT GGCTCGCGGA CCACGCCGTC CTGGGAGCG TCCTGCTCC 12600
 L P D R D G L L L T G R L S L R T H P W L A D H A V L G S V L L P

CGGCGCGCG ATGTCGAAC TCGCCGCGCA CGCTGCGGAG TCCGCGGTC TGCTGACGT GCGGAGCTG ACCCTCTTG AACCGCTGT ACTGCCCGAG 12700
 G A A M V E L A A H A A E S A G L R D V R E L T L L E P L V L P E

CACGTGGCG TCGAGCTGCG CGTGACGGTC GGGCGCGCG CCGGAGAGCC CGGTGGCGAG TCGGCGGCGG ACGGCGCAG GCCGCTGCC CTCCACTCGC 12800
 H G G V E L R V T V G A P A G E P G G E S A G D G A R P V S L H S R

▼
 Acc65I KpnI MscI
 ▼
 Bali

GGCTCGCGA CGGCGCGCG GGTACCGCCT GGTCTGCCA CGGACCGGT CTGTGGCCA CCGACCGCC CGAGCTTCCC GTGCGCGCGG ACGTGCGGC 12900
 L A D A P A G T A W S C H A T G L L A T D R P E L P V A P D R A A

CATGTGGCG CCGCAGGCG CCGAGGAGT GCCGCTCGAC GGTCTTACG AGCGCTCGA CCGGAACGGC CTCGCTTCG GTCCGCTGT CCAGGGGCTG 13000
 M W P P Q G A E E V P L D G L Y E R L D G N G L A F G P L F Q G L

FIG. 31Z

67/164

plkPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890

BsmI
▼

AACGCGGTGT GCGGTACGA GGTGAGTTC TTCGCCGACA TCGCGTCCC CGCCACCACG AATGCGACG CGCCCGCGAC CGGAACGGC GCGGGAGTG 13100
N A V W R Y E G E V F A D I A L P A T T N A T A P A T A N G G G S A

CGGCGGCGG CCCCTACGGC ATCCACCCCG CCTGCTCGA CGTTGCTG CACGCCATCG CGGTGCGCG TCTGTCGAC GAGCCCGAGC TCGTCGCGT 13200
A A A P Y G I H P A L L D A S L H A I A V G G L V D E P E L V R V

ApalI
▼

CCCCTTCAC TGGAGCGTG TCACCGTGCA CGCGGCGGT GCCGGGTCCG TCTGCGCTCC GCGGGGACG ACGCGTCTC GCTGTCCCTG 13300
P F H W S G V T V H A A G A A A A R V R L A S A G T D A V S L S L

ACGGACGGCG AGGACGCCC GCTGGTCTCC GTGGAACGGC TCACGCTGCG CCGGTACCG GCCGATCAG CGCGGCGAG CCGGTGCGC GGGCTGATG 13400
T D G E G R P L V S V E R L T L R P V T A D Q A A A S R V G G L M H

ACCGGTGCG CTGGCGTCG TACGCCCTCG CCTGTCGCG CGAACAGGAC CCGCAGCCA CTTGCTACG GCCGACCGC GTCTCGGCA AGGACGAGT 13500
R V A W R P Y A L A S S G E Q D P H A T S Y G P T A V L G K D E L

FIG. 31AA

68/164

plkPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890

GAAGTCGCC GCGGCCCTGG AGTCCGCGGG CGTCGAAGTC GGGCTCTACC CCGACTTGGC CCGGCTGTCC CAGGACGTGG CCGCCCGCGC CCGCGCGCCC 13600
 K V A A A L E S A G V E V G L Y P D L A A L S Q D V A A G A P A P

CGTACCGTCC TTGGCGCGCT GCCCGCGGGT CCGCGCGGAG GCGTGTACGG GGCACGCTGG CCGGACGCT GAGCTGCTC CAGGCTGCG 13700
 R T V L A P L P A G P A D G G A E G V R G T V A R T L E L L Q A W L

TGGCCGACGA GCACCTCGCG GGCACCGCGG TGCTCTGGT CACCCGCGGT GCGGTGCGGG ACCCGAGGG GTCCGCGCGC GACGATGGCG GCGAGGACCT 13800
 A D E H L A G T R L L L V T R G A V R D P E G S G A D D G G E D L

NotI

GTCCGACGCG GCGCGCTGGG GTCTCGTACG GACCGCGCAG ACCGAGACCC CCGGCCGCTT CCGCCTTCTC GACCTGGCGG ACGACGCTC GTCGTACCGG 13900
 S H A A A W G L V R T A Q T E N P G R F G L L D L A D D A S S Y R

BstXI

ACCCCTGCCGT CCGTGCTCTC CGACGCGGGC CTGCGCGGACG AACCGCAGCT CGCCCTGCAC GACGGCACCA TCAGGCTGGC CCGCCTGGC TCCGTCCGGC 14000
 T L P S V L S D A G L R D E P Q L A L H D G T I R L A R L A S V R P

FIG. 31BB

69/164

plkPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890

CCGAGACCGG CACCGCGCA CCGCGCTCG CCCGGAGGG CACGGTCCTG CTGACCGCG GCACCGGCG CCTGGGCGA CTGGTCGCC GGCACGTGGT 14100
E T G T A A P A L A P E G T V L L T G G T G G L G G L V A R H V V

PmlI
▼

ApaLI
▼

GGGCGAGTGG GGCCTACGAC GCCTGCTGCT GGTGAGCCG CGGGGCACGG AGCCCCCGG CGCCGACGAG CTCGTGCACG AGCTGGAGGC CCTGGGAGCC 14200
G E W G V R R L L L V S R R G T D A P G A D E L V H E L E A L G A

GACGTCTCGG TGGCCGCGTG CGACGTCGCC GACCGCGAAG CCTCACCGC CGTACTCGAC GCCATCCCCG CCGAACACCC GCTCACCGCG GTCGTCCACA 14300
D V S V A A C D V A D R E A L T A V L D A I P A E H P L T A V V H T

CGGCAGGCGT CCTCTCCGAC GGCACCTCC CGTCCATGAC GACGGAGGAC GTGGAACACG TACTGCGGCC CAAGGTCGAC GCCGCGTTCC TCCTCGACGA 14400
A G V L S D G T L P S M T T E D V E H V L R P K V D A A F L L D E

ACTACCTCG AGCCCCGCAAT ACGACCTGGC AGGTTGTC ATGTTCTCCT CCGCCCGCGC CGTCTTCGGT GGCGCGGGGC AGGGCGCCTA CGCCGCCGCC 14500
L T S T P A Y D L A A F V M F S S A A A V F G G A G Q G A Y A A A

FIG. 31CC

70/164

pkPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890

AAGCCACCC TCGACGCCCT CGCCTGGCGC CGCCGGGCAG CCGGACTCCC CGCCCTCTCC CTCGGCTGGG GCCTCTGGC CGAGACCAGC GGCATGACCG 14600
N A T L D A L A W R R R A A G L P A L S L G W G L W A E T S G M T G

GGAGCTCGG CCAGCGGAC CTGCGCCGGA TGAGCCGCGC GGGCATCGGC GGGATCAGCG ACGCCGAGGG CATCGCGCTC CTCGACGCCG CCCTCCGCGA 14700
E L G Q A D L R R M S R A G I G G I S D A E G I A L L D A A L R D

CGACCGCCAC CCGGTCTGTC TGCCCTGCG GCTCGACGCC GCCGGGCTGC GGGACGCGC CCGGCCGGA TCCCGGCGCT CTTCGGGAC 14800
D R H P V L L P L R L D A A G L R D A A G N D P A G I P A L F R D

GTCTCGGCG CCAGGACCGT CCGGGCCCGG CCGTCCGCGG CCTCGCCTC GACGACAGCC GGGACGCGC GCACGCCGGG GACGCGGAC GGC CGCGCGG 14900
V V G A R T V R A R P S A A S T T A G T A G T P G T A D G A A E

XhoI
PaeR7I



AAACGCGGC GGTACGCTC GCCGACCGG CCGCCACCGT GGACGGGCC GCACGGCAGC GCCTGCTGCT CGAGTTCGTC GTCGCGGAGG TCGCCGAAGT 15000
T A A V T L A D R A A T V D G P A R Q R L L L E F V V G E V A E V

FIG. 31DD

71/164

pikPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890

Seal

▼
 ACTCGGCCAC GCCCGCGGTC ACGGATCGA CGCCGAACGG GGTTCCTCG ACCTCGGCTT CGACTCCCTG ACCGCCGTCG AACTCCGCAA CCGGTCAAC 15100
 L G H A R G H R I D A E R G F L D L G F D S L T A V E L R N R L N

TCCGCGGTC GCCTGCGCT CCGGGGACC CTGGTCTTCG ACCACCCAAG CCGGGGGCA CTGGCTCC ACCTGGACGC CGAGCTGCCG CCGGCGCCT 15200
 S A G G L A L P A T L V F D H P S P A A L A S H L D A E L P R G A S

CGGACGAGGA CGGAGCCGGG AACCGGAAC GGAACGAGAA CCGGACGACG GCGTCCCGA GCACCGCGG GACGGACGC CTGCTGGCAC AACTGACCCG 15300
 D Q D G A G N R N G N E N G T T A S R S T A E T D A L L A Q L T R

CCTGGAGGC GCCTTGGTC TGACGGCCT CTGGACGCC CCGGGAGCG AAGAAGTCTT GGAGCACCTG CCGTCCCTGC GCTCGATGGT CACGGCGGAG 15400
 L E G A L V L T G L S D A P G S E E V L E H L R S L R S M V T G E

ACCGGAGCG GGACCGGTC CGGAGCCCG GTCCGGCCG GACGGGCGG CGAGGACCG CCCTGGGCG CCGGGGACGG AGCCGGGGGC GGGAGTGAGG 15500
 T G T G T A S G A P D G A G S G A E D R P W A A G D G A G G G S E D

FIG. 31EE

72/164

plkPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890

BspHI

ACGGCGCGG AGTGC CGGAC CCTCGGCGG GGAATCTTTC GGCTCTCTCG ACCAGGACCC CAGCACGGAC TGATCCCTGC CGCACGGTGC 15600
 G A G V P D F M N A S A E E L F G L L D Q D P S T D

XmnI

CCTCCCGCCC CGGACCCCGT CCGGGGCACC TCGACTCGAA TCACTTCATG CGCGCCTCGG GCGCTCCAG GAACTCAGG GGAAGCGTG TCCACGGTGA 15700
 V S T V N

ACGAAGAGAA GTACCTCGAC TACCTGCGTC GTGCCACGGC GGACTCTCCAC GAGGCCCGTG GCCGCCTCCG CGAGCTGGAG GCGAAGGCGG GCGAGCCCGT 15800
 E E K Y L D Y L R R A T A D L H E A R G R L R E L E A K A G E P V

GGCGATCGTC GGCATGGCCT GCGCCTGCCC CGGCGGCGTC GCCTCGCCCG AGGACCTGTG GCGGCTGGTG GCCGGCGCG AGGACGCGAT CTCGGAGTTC 15900
 A I V G M A C R L P G G V A S P E D L W R L V A G G E D A I S E F

CCCCAGGACC GCGGCTGGGA CGTGGAGGGC CTGTACGACC CGAACCCGGA GGCCACGGGC AAGAGTTACG CCCGGGAGGC CGGATTCCTG TACGAGGCGG 16000
 P Q D R G W D V E G L Y D P N P E A T G K S Y A R E A G F L Y E A G

FIG. 31FF

73/164

plkPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890

GCGAGTTTGA CGCCGACTTC TTGGGATCT CGCCGCGGGA GGCCCTCGCC ATGGACCCGC AGCAGGTCTT CCTCCTGGAG GCCTCCTGGG AGCGTTTGA 16100
E F D A D F F G I S P R E A L A M D P Q Q R L L L E A S W E A F E

BamHI



GCACGCGGG ATCCCGGGG CCACCGGCG CGGCACCTCG GTGCGGTCT TCACCGCGCT GATGTACCAC GACTACGCCA CCCGTCTCAC CGATGTCCCG 16200
H A G I P A A T A R G T S V G V F T G V M Y H D Y A T R L T D V P

GAGGCATCG AGGGTACCT GGGCACGGC AACTCCGGCA GTGTGCTC GTGCGGCTC GCGTACACGC TTGGCCTGGA GGGGCCGGCC GTCACGGTCG 16300
E G I E G Y L G T G N S G S V A S G R V A Y T L G L E G P A V T V D

ACACCGCCTG CTCGTCCTCG CTGCTGCCC TGCACCTCG CGTGCAGGCC CTGCGCAAGG GCGAGGTGGA CATGGCGCTC GCCGCGGCG TGACGGTCAT 16400
T A C S S S L V A L H L A V Q A L R K G E V D M A L A G G V T V M

XmnI



GTGACGCCC AGCACCTTCG TCGAGTTTCAG CCGTCAGCG GGGCTGGCG CGGACGGCG GTCGAAGTCC TTCTCGTTCGA CGGCCGACGG CACCAGCTGG 16500
S T P S T F V E F S R Q R G L A P D G R S K S F S S T A D G T S W

FIG. 31GG

74/164

plkPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890

BamHI
▼

TCCGAGGGCG TCGGCGTCCT CCTCGTCGAG CGCCTGTCCG ACGGCGTGG CAAGGGCCAT CGGATCCTCG CCGTGGTCCG GGGCACCGCC GTCAACCCAGG 16600
 S E G V G V L L V E R L S D A R R K G H R I L A V V R G T A V N Q D

ACGGGGCCAG CAGGGGCTC ACGGCTCGA ACGGGCCGTC GCAGCAGCGC GTCATCCGAC GTGCCCTGGC GGACGCCCGG CTCACGACCT CCGACGTGGA 16700
 G A S S G L T A P N G P S Q Q R V I R R A L A D A R L T T S D V D

CGTCGTGCGAG GCCACGGCA CGGTACGG ACTCGGGCGAC CCGATCGAGG CGCAGGCGGT CATCGCCACG TACGGGCGG GCCGTGACGG CGAACAGCCG 16800
 V V E A H G T G T R L G D P I E A Q A V I A T Y G Q G R D G E Q P

BclI
▼

CTGCGCCTCG GGTGTTGAA GTCCAACATC GGACACACCC AGGCCGCGC CGGTGTCTCC GCGGTGATCA AGATGGTCCA GCGCATGCGC CACGGCGTCC 16900
 L R L G S L K S N I G H T Q A A A G V S G V I K M V Q A M R H G V L

PmlI
▼

TGCCGAGAC GCTCCACGTG GAGAAGCCGA CCGACCCAGGT GGACTGGTCC GCGGGGCGG TCGAGCTGCT CACCGAGGCC ATGGAAGTGC CGGACAGGG 17000
 P K T L H V E K P T D Q V D W S A G A V E L L T E A M D W P D K G

FIG. 31HH

75/164

plkPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890

CGACGGCGGA CTGCGCAGGG CCGCGGTCTC CTCCTTCGGC GTCAGCGGA CGAACGGCA CGTCGTGCTC GAAGAGGCC CCGCGGCCGA GGAGACCCCT 17100
D G G L R R A A V S S F G V S G T N A H V V L E E A P A A E E T P

GCCTCCGAGG CGACCCCGGC CGTCGAGCCG TCGGTCCGGC CCGGCTGGT GCCGTGGCTG GTGTCCGGA AGACTCCGC CCGCTGGAC GCCAGATCG 17200
A S E A T P A V E P S V G A G L V P W L V S A K T P A A L D A Q I G

AscI
▼

GACGCTCGC CCGGTTCCG TCGCAGGCG GTACGGACGC CGCCGATCCG GCGCGGTGCTG CTCGCTACT GGCCGGCGGG CCGCCGAGT TCGAGCACCG 17300
R L A A F A S Q G R T D A A D P G A V A R V L A G G R A E F E H R

GGCGTCGTG CTCGGCACCG GACAGGACGA TTTCCGCGCAG GCGCTGACCG CTCGGAAGG ACTGATACGC GGCACGCCCT CCGACGTGG CCGGTTGGCG 17400
A V V L G T G Q D D F A Q A L T A P E G L I R G T P S D V G R V A

XmnI
▼

TTCTGTTC CCGGTCAGGG CACGACGTGG GCCGGGATGG GCGCCGAACT CCTCGACGTG TCGAAGAGT TCGCGCGGC CATGGCCGAG TCGAGAGCG 17500
F V F P G Q G T Q W A G M G A E L L D V S K E F A A A M A E C E S A

FIG. 31II

76/164

plPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890

ASCI
▼

CGCTCTCCG CTATGTCGAC TGGTCGCTGG AGCCGTCGT CCGGCAGGCG CCGGCGCGC CCACGCTGGA GCGGTCGAC GTCGTCCAGC CCGTGACCTT 17600
L S R Y V D W S L E A V V R Q A P G A P T L E R V D V V Q P V T F

CGCTGTCATG GTTTCGCTGG CGAAGGTCTG GCAGCACCAC GCGGTGACGC CGCAGGCCGT CGTCGGCCAC TCGCAGGCG AGATCGCCGC CCGGTACGTC 17700
A V M V S L A K V W Q H H G V T P Q A V V G H S Q G E I A A A Y V

GCGGTGCC TCACCTCGA CGACGCCGCC CCGTCTGTC CAATCCATC GCCGCCACC TCGCCGCGAA GCGCGGCATG ATCTCCCTCG 17800
A G A L T L D D A A R V V T L R S K S I A A H L A G K G G M I S L A

CCCTCAGCGA GGAAGCCACC CCGCAGCGCA TCGAGAACCT CCACGACTG TCGATCGCCG CCGTCAACGG CCCACCGCC ACCGTGGTTT CCGGCGACCC 17900
L S E E A T R Q R I E N L H G L S I A A V N G P T A T V V S G D P

CACCCAGATC CAAGAGCTG CTCAGGCGTG TGAGGCCGAC GGGGTCCGG CACGGATCAT CCGGTCGAC TAGCCCTCCC ACAGCGCCCA CCGTCGAGACC 18000
T Q I Q E L A Q A C E A D G V R A R I I P V D Y A S H S A H V E T

FIG. 31JJ

77/164

plPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890

EspI
 Bpu1102I
 Bsu36I

ATCGAGAGCG AACTCGCGGA GGTCTCTGCC GGGCTCAGCC CGCGACACCC TGAGTGCCG TTCTTCTCGA CACTCGAAGG CGCTGGATC ACCGAGCCG 18100
 I E S E L A E V L A G L S P R T P E V P F F S T L E G A W I T E P V

KpnI
 Acc65I

TGCTCGAGCG CACTTACTGG TACCGCAACC TCCGCCACCG CGTCGGCTTC GCCCGGCCG TCGAGACCCCT CGCCACCGAC GAAGCTTCA CCCACTTCAT 18200
 L D G T Y W Y R N L R H R V G F A P A V E T L A T D E G F T H F I

CGAGGTGAGC GCCACCCCG TCCTCACCAT GACCTCTCCC GAGACCGTCA CCGGCTCCG CACCTCCGC CGCGAACAGG GAGGCCAGGA GCGTCTGGTC 18300
 E V S A H P V L T M T L P E T V T G L G T L R R E Q G G Q E R L V

ACCTCACTCG CGAAGCCTG GACCAACGGC CTCACCATCG ACTGGGCGCC CGTCTCTCCC ACCGCAACCG GCCACACCC CGAGCTCCCC ACCTACGCT 18400
 T S L A E A W T N G L T I D W A P V L P T A T G H H P E L P T Y A F

TCCAGCGCGG TCACTACTGG CTCCACGACT CCCCAGGCTCC CCAGGGCTCC GTGCAGGACT CCTGGCGCTA CCGCATCGAC TGAAGCGCC TCGCGGTCCG 18500
 Q R R H Y W L H D S P A V Q G S V Q D S W R Y R I D W K R L A V A

FIG. 31KK

78/164

pkPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890

MluI



CGACGCGTCC GAGCGCGCCG GGCTGTCCGG GCGCTGGCTC GTCGTGCTCC CCGAGGACCG TTCCGCCGAG GCCGCCCGG TGTGCGCCG GCTGTCCGCG 18600
 D A S E R A G L S G R W L V V P E D R S A E A A P V L A A L S G

GCCGCGCCG ACCCGGTACA GCTGGACGTG TCCCGCTGG GCGACCGGCA GCGGCTGCGC GCGACGCTGG GCGAGGCCCT GCGGCGCGCC GGTGGAGCCG 18700
 A G A D P V Q L D V S P L G D R Q R L A A T L G E A L A A G G A V

TGACCGCGT CCTCTCGTG CTCGCGTGGG ACGAGAGCGC GCACCGCGGC CACCCCGCC CTTACACCG GGGCACCGGC GCCACCTCA CCCTGTRCA 18800
 D G V L S L L A W D E S A H P G H P A P F T R G T G A T L T L V Q

GGCGCTGGAG GACGCGCGCG TCGCGCGCC GCTGTGTTGC GTGACCCACG GCGCGGTGC CGTGGCGCGG GCGACCAAG TCACCTCCCC CGCCCGGCC 18900
 A L E D A G V A A P L W C V T H G A V S V G R A D H V T S P A Q A

ATGCTGTGG GCATGGGCG GGTCGCGCGC CTGGAGCACC CCGAGCGGTG GGGCGSCCTG ATCGACCTGC CCTGGGAGCG CGACCGGGCG GCCCTGGACC 19000
 M V W G M G R V A A L E H P E R W G G L I D L P S D A D R A A L D R

FIG. 31LL

79/164

plkPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890

GCATGACCAC GGTCTCTGCC GGGCGTACGG GTGAGGACCA GGTGCGGTA CGCGCTCCG GGCTGCTCG CGCGGCCTC GTCCGGCCT CCCTCCCGC 19100
M T T V L A G G T G E D Q V A V R A S G L L A R R L V R A S L P A

NotI



GCACGGCAG GCTTCGCCGT GGTGGCAGGC CGACGGCAG GTGCTCGTCA CCGTGCCGA GGAGCCTGCG GCCGCCGAG CCGCAGCCG GCTGGCCCG 19200
H G T A S P W Q A D G T V L V T G A E E P A A A E A A R R L A R

GACGGCGCG GACACCTCCT CCTCCACACC ACCCCTCCG GCAGCGAAG CGCCGAGGC ACCTCCGCTG CCGCCGAGGA CTCGGGCCTC GCCGGGCTCG 19300
D G A G H L L L H T T P S G S E G A E G T S G A A E D S G L A G L V

NotI



TCGCCGAAC CGCGGACCTG GGGCGGACGG CCACCGTCGT GACCTGGGAC CTCACGGACG CGGAGGCGGC CGCCGGGCTG CTCGGCCGCG TCTCCGACGC 19400
A E L A D L G A T A T V V T C D L T D A E A A A R L L A G V S D A

EspI

Bpu1102I



GCACCCGCTC AGCGCGCTCC TCCACCTGCC GCCACCGTC GACTCGGAGC CGCTCGCGC GACCGACGCG GACCGGCTCG CCCGTGTGCT GACCGCGAAG 19500
H P L S A V L H L P P T V D S E P L A A T D A D A L A R V V T A K

FIG. 31MM

80/164

pikPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890

GCACCGCCG CGCTCCACCT GGACCGCCTC CTGCGGAGG CCGCGGCTGC CGAGGCCGT CGGCCCGTCC TGGTCCTCTT CTCCTCGGTC GCCGCGATCT 19600
 A T A A L H L D R L L R E A A A A G G R P P V L V L F S S V A A I W

GGGGCGGCG CGGTACGCC GCGTACGGC CTTCTCGAC GCCCTGCCG GTCAGCACCG GGCCGACGCG CCCACCGTGA CCTCGGTGGC 19700
 G G A G Q G A Y A A G T A F L D A L A G Q H R A D G P T V T S V A

CTGAGCCCC TGGAGGGCA GCCGCGTAC CGAGGGTGG ACCGGGAGC GGCTGCCCG CTGCGCCTG CGCCCGCTCG GCGGCTCACC 19800
 W S P W E G S R V T E G A T G E R L R R L G L R P L A P A T A L T

GCCCTGGACA CCGCGCTCG CCACGGGAC ACCGCGGTCA CGATCGCGA CGTCGACTGG TCGAGCTTCG CCCCCGGCTT CACCACGGCC CGGCCGGGCA 19900
 A L D T A L G H G D T A V T I A D V D W S S F A P G F T T A R P G T

Asci

CCCTCCTCGC CGATCTGCCC GAGGCGGCC CCGCGTCTGA CGAGCAGCAG TCGACGACG CCGCCGACGA CACCGTCTG AGCCGCGAGC TCGGTGCGCT 20000
 L L A D L P E A R R A L D E Q Q S T T A A D D T V L S R E L G A L

FIG. 31NN

81/164

pkPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890

CACCGGCGCC GAACAGCAGC GCCGTATGCA GGAGTTGGTC CGCGAGCACC TCGCCGTGGT CCTCAACCAC CCCTCCCCCG AGCCCGTCGA CACGGGGCGG 20100
 T G A E Q Q R R M Q E L V R E H L A V V L N H P S P E A V D T G R

BstXI
 ▼

GCCTTCGGTG ACCTCGGATT CGACTCGCTG ACGCGGGTCG AGCTCCGCAA CCGCCTCAAG AACGCCACCG GCCTGGCCCT CCGGGCCACT CTGGTCTTCG 20200
 A F R D L G F D S L T A V E L R N R L K N A T G L A L P A T L V F D

ACTACCCGAC CCCCCGGAGC CTGGCGGAGT TCCTCCTCGC GGAGATCCTG GCGAGCAGG CCGGTGCCG CGAGCAGCTT CCGGTGGACG GCGGGGTGGA 20300
 Y P T P R T L A E F L L A E I L G E Q A G A G E Q L P V D G G V D

CGACGAGCCC GTCGCGATCG TCGGCATGGC GTGCCGCTCG CCGGGCGGTG TCGCCTCGCC GGAGGACCTG TGGCGGCTGG TGGCCGCGCG CGAGGACGCG 20400
 D E P V A I V G M A C R L P G G V A S P E D L W R L V A G G E D A

MluI
 ▼

ATCTCCGGCT TCCCGCAGGA CCGCGGCTGG GACGTGGAGG GGCTGTACGA CCGGACCCG GACGCGTCCG GCGGACGTA CTGCCGTGCC GGTGGCTTCC 20500
 I S G F P Q D R G W D V E G L Y D P D P D A S G R T Y C R A G G F L

FIG. 3100

82/164

pkPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890

TCGACGAGGC GGGCGAGTTC GACGCCGACT TCTTCGGGAT CTGCGCGCGC GAGGCCCTCG CCATGGACCC GCAGCAGCGG CTCCTCTCTG AGACCTCCTG 20600
 D E A G E F D A D F F G I S P R E A L A M D P Q Q R L L L E T S W

GGAGGCGGTC GAGGACGCCG GGATCGACCC GACCTCCCTT CAGGGGCGAG AGGTGCGCGT GTTCGCGGCG ACCAAGCGCC CCCACTACGA GCCGCTGCTC 20700
 E A V E D A G I D P T S L Q G Q Q V G V F A G T N G P H Y E P L L

CGCAACACCG CCGAGGATCT TGAGGGTTAC GTCGGGACCG GCAACGCCG CAGCATCATG TCGGGCGGTG TCTCGTACAC CCTCGGCGCTG GAGGCGCGG 20800
 R N T A E D L E G Y V G T G N A A S I M S G R V S Y T L G L E G P A

BsmI

CCGTCACGGT CGACACCGCC TGCTCCTCCT CGCTGCTGCG CCTGCACCTC GCCGTGCGAG CCCTGCGCAA GGGCGAATGC GGACTGGCGC TCGCGGCGG 20900
 V T V D T A C S S L V A L H L A V Q A L R K G E C G L A L A G G

XmnI

TGTGACGGTC ATGTCGACGC CCACGACGTT CGTGGAGTTC AGCCGGCAGC GCGGGCTCGC GGAGGACGGC CGGTGGAAGG CGTTGCGCGC GTCGCGGAC 21000
 V T V M S T P T T F V E F S R Q R G L A E D G R S K A F A S A D

FIG. 31PP

83/164

plkPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890

SphI

GGCTTCGGCC CGGCGGAGGG CGTCGGCATG CTCCTCGTCG AGCGCCTGTC GGACGCCCGC CGCAACGGAC ACCGTGTGCT GCGGTCGTG CGCGGCAGCG 21100
 G F G P A E G V G M L L V E R L S D A R R N G H R V L A V V R G S A

CGGTCAACCA GGACGGCGG AGCAACGGCC TGACCGGCC TCGCAGCAGC GCGTCATCG GCGCGGCTC GCGGACGCC GACTGACGAC 21200
 V N Q D G A S N G L T A P N G P S Q Q R V I R R A L A D A R L T T

CGCGACGTC GACGTCGTC AGGCCACGG CACGGGCGCG CGACTCGCG ACCCGATCGA GGCACAGGCC CTCATCGCCA CCTACGGCCA GGGGCGCGAC 21300
 A D V D V V E A H G T G T R L G D P I E A Q A L I A T Y G Q G R D

ACCGAACAGC CGCTGGCCT GGGTCGTTG AAGTCCAACA TCGGACACAC CCAGGCGCC GCGGTGTCT CCGGCATCAT CAAGATGGTC CAGGCGATGC 21400
 T E Q P L R L G S L K S N I G H T Q A A A G V S G I I K M V Q A M R

PmlI

GCCACGGCGT CCTGCCGAG ACGTCCACG TGGACCGGCC GTCGGACCAG ATCGACTGGT CGGCGGGCAC GGTGAGCTG CTCACCGAGG CCATGGACTG 21500
 H G V L P K T L H V D R P S D Q I D W S A G T V E L L T E A M D W

FIG. 31QQ

84/164

plkPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890
GCCGAGGAAG	CAGGAGGGCG	GGCTGCGCCG	CGCGGCGGTC	TCCTCCTTCG	GCATCAGCGG	CACGAACGCG	CACATCGTGC	TGGAAGAAGC	CCCGTCGAC
P R K Q E G G L R R A A V S S F G I S G T N A H I V L E E A P V D									
GAGGACGCCC	CGGCGGACGA	GCCGTGCGTC	GGCGGTGTGG	TGCCGTGGCT	CGTGTCCGCG	AAGACTCCGG	CCGCGCTGGA	CGCCAGATC	GGACGCTCG
E D A P A D E P S V G G V V P W L V S A K T P A A L D A Q I G R L A									
CCGCGTTGCG	CTCGCAGGCG	CGTACGGACG	CCGCGGATCC	GGGCGCGGTC	GCTCGGTAC	TGGCCGCGCG	GCGTGCGCAG	TTCGAGCACC	GGGCGCTCG
A F A S Q G R T D A A D P G A V A R V L A G G R A Q F E H R A V A									
		NotI		Bsu36I					
GCTCGGACCC	GGACAGGACG	ACCTGCGCGC	CGCACTGGCC	GGCGCTGAGG	GTCGTGTCGG	GGGTGTGGCC	TCCGTGTGG	GTCGAGTGGC	GTTCTGTGTC
L G T G Q D D L A A A L A A P E G L V R G V A S G V G R V A F V F									
								XmnI	
CCGGACAGG	GCACGAGTG	GGCCGCGGATG	GGTGCCGAAC	TCCTCGACGT	GTCGAAGGAG	TTCGCGGCGG	CCATGGCCGA	GTGCGAGGCC	GCGCTCGCTC
P G Q G T Q W A G M G A E L L D V S K E F A A A M A E C E A A L A P									

FIG. 31RR

85/164

plkPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890

ASCI
▼

CGTACGTGGA CTGGTCGCTG GAGGCCGTCG TCCGACAGGC CCCCGGCGCG CCCACGCTGG AGCGGTCGA TGTCGTCCAG CCGGTGACGT TCGCCGTGAT 22100
Y V D W S L E A V V R Q A P G A P T L E R V D V V Q P V T F A V M

GGTCTCGCTG GCGAAGGTCT GGCAGCACCA CGGGGTGACC CCGCAAGCCG TCGTCGGCCA CTCGACGGGC GAGATCGCG CCGGTGACGT CGCCGGTGCC 22200
V S L A K V W Q H H G V T P Q A V V G H S Q G E I A A A Y V A G A

EspI
Bpu1102I
▼

SphI
▼

CTGAGCCTGG ACGACGCCGC TCGTGTCGTG ACCCTGGCA GCAAGTCCAT CGGCGCCAC CTCGCGGGCC AGGGCGGCAT GCTGTCCCTC GCGCTGAGCG 22300
L S L D D A A R V V T L R S K S I G A H L A G Q G G M L S L A L S E

AGGCGGCCGT TGTGAGCGA CTGGCCGGGT TCGACGGGCT GTCCGTGCGC GCCGTCAACG GGCCTACGC CACCGTGGTT TCGGGCGACC CGACCCAGAT 22400
A A V V E R L A G F D G L S V A A V N G P T A T V V S G D P T Q I

CCAAGAGCTC GCTCAGGCGT GTGAGGCCGA CGGGGTCCGC GCAAGGATCA TCCCGGTGCA CTACGCCTCC CACAGCGCC ACGTCGAGAC CATCGAGAGC 22500
Q E L A Q A C E A D G V R A R I I P V D Y A S H S A H V E T I E S

FIG. 31SS

86/164

plkPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890

GAAGTCGGCG AGTCCTCTGGC GGGGTGTGTC CCCCAGACAC CCCAGGTCCC CTCTCTCTCC ACCCTCGAAG GCGCCTGGAT CACCGAACCC GCCTCGACG 22600
 E L A D V L A G L S P Q T P Q V P Q F F S T L E G A W I T E P A L D G

KpnI
 Acc65I

MscI
 BalI

GGGGTACTG GTACCGCAAC CTCGCCCATC GTGTGGGCTT CGCCCCGGCC GTCGAATCCC TGGCCACCGA CGAAGGCTTC ACCCACTTCG TCGAGGTGAG 22700
 G Y W Y R N L R H R V G F A P A V E T L A T D E G F T H F V E V S

CGCCCCACCC GTCCTCACCA TGGCGGTGCC CGAGACCGTC ACCGACTCG GCACCTCCG CCGTGACAAC GCGGACAGC ACCGCTTCAC CACCTCCCTC 22800
 A H P V L T M A L P E T V T G L G T L R R D N G G Q H R L T T S L

GCCGAGGCT GGGCCAACGG CCTCACCCTC GACTGGGCT CTCTCTCTCC CACCAGACC ACCCACCCCG ATCTGCCAC CTACGCCTTC CAGACCGAGC 22900
 A E A W A N G L T V D W A S L L P T T T T H P D L P T Y A F Q T E R

Eco47III

GCTACTGGCC GCAGCCGAC CTCTCCGCG CCGGTGACAT CACTCCGCC GGTCTCGGG CGGCCGAGCA CCCGCTGCTC GCGCGGCGG TGGCGCTCGC 23000
 Y W P Q P D L S A A G D I T S A G L G A A E H P L L G A A V A L A

FIG. 31TT

87/164

plPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890
GGACTCCGAC	GGTGCCTGC	TCACGGGGAG	CCTCTCCCTC	CGTACGCACC	CCTGGCTGGC	GGACCACGCG	GTGGCCGGCA	CCGTGCTGCT	GCCGGGAACG
D S D	G C L L	T G S L	S L S L	R T H P	W L A D	H A V A	G T V L	L L P G	T
23100									
GGTTTCGTGG	AGCTGGCGTT	CCGAGCCGGG	GACCAGGTGC	GTTGCGATCT	GGTCGAGGAG	CTCACCTCTC	ACGCGCGCT	CGTCTGCCC	CGTCGTGGCG
A F V	E L A F	R A G D	Q V G C	D L V E	L T L D	A P L V	L P R R	G A	
23200									
CGGTCCGTGT	GCAGCTGCC	GTCCGGCGGA	GCGACGAGTC	CGGGCGTGT	ACCTTCGGGC	TCTACGCGCA	CCCGGAGGAC	GCGCCGGGCG	AGCGGAGTG
V R V	Q L S V	G A S D	E S G R	R T F G	L Y A H	P E D A	P G E A	E W	
23300									
GACGCGCAC	GCCACCGGTG	TGCTGGCCGC	CCGTGCGGAC	CGCACCGCCC	CCGTGCGCCA	CCCGGAGGCC	TGGCCGCCGC	CGGGCGCCGA	GCCGGTGGAC
T R H	A T G V	L A A R	A D R A	D R T A	P V A D	P E A W	P P P G	A E P V	D
23400									
GTGACGGTC	TGTACGAGCG	CTTCGGCGCG	AACGGCTACG	GCTACGGCCC	CCTCTTCCAG	GCGTCCGTG	GTGTCTGGCG	GCGTGGCGAC	GAGGTGTTCG
V D G	L Y E R	F A A N	G Y G Y	G P L F	Q G V R	G V W R	R R G D	E V F A	
23500									

Eco47III

FIG. 31UU

88/164

plkPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890
CCGACGTGGC	CCTGCCGGCC	GAGGTGCGCG	GTGCCGAGGG	CGCGCGGTTC	GGCCTTCACC	CGCGCTGCT	CGACGCCGCC	GTGCAGGCGG	CCGCTGCGGG
D V A	L P A	E V A	G A E	G A R	F G L	H P A	L L D	A A V	Q A A
G A G									
CGGGGCGTT	CGGCGCGGC	ACGCGGCTGC	CGTTGCGCTG	GAGCGGGATC	TCCTGTACGC	GGTCGGCGCC	ACCGCCTTCC	GCGTGGGCT	GGCCCCCGCC
R G V	R R G	H A A	A V R	L E R	D L L	Y A V	G A T	A L R	V R L
A A									
GGCCCGGACA	CGGTGTCCGT	GAGCGCGCC	GACTCCCTCC	GCGCGCGGT	GTTGCGCGCG	GACTCCCTCA	CGGTGCTGCC	CGTCGACCCC	GCGCAGCTGG
G P D	T V S	V S A	A D S	S G Q	P V F	A A D	S L T	V L P	V D P
A A									
CGGCCTTCAG	CGACCCGACT	CTGACCGCGC	TGCACCTGCT	TGCACCTGCT	GGAGTGGACC	GCCTGGGACG	GTGCCGCGCA	GGCCCTTGCCC	GGCGGGTGC
A F S	D P T	L D A	L H L	L L E	W T A	W D G	A A Q	A L P	G A V
L L									
CGACGCGGAC	GGTCTGCGCG	CGGCGGTGCG	CGCCGCTGCC	ACGAGGTCC	TGTCCTTCCC	GGACCTTACG	GACCTGTGG	AGGCCGTGA	CCGGGGCGGAG
D A D	G L A	A A L	R A G	G T E	V L S	F P D	L T D	L V E	A V D

FIG. 31W

89/164

plkPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890
ACCCCGGCC	CCTGTGGCC	TGCCCCGCC	CGGCCCCGA	TGGCCCGAG	CATGTCGGG	AGGCCCTGCA	CGGTCGCTC	GCGCTGATC	24100
T P A P	A T V L V A	C P A A	G P D G P E	H V R E	A L H G S L	A L M Q			
AGGCTTGGCT	GGCCGACGAG	CGGTTACCG	ATGGGCGCCT	GGTGCTCGTG	ACCCGCGACG	CGGTCGCGC	CCGTTCCGGC	GACGGCCTGC	24200
A W L A D E	R F T D G R L	V L V T R D A	V A A R S G	D G L R S T G					
ACAGGCGGCC	GTCTGGGGCC	TCGGCCGGTC	CGCGCAGACG	GAGAGCCCGG	GCCGGTTCTGT	CCTGCTCGAC	CTCGCCGGGG	AAGCCCGGAC	24300
Q A A V W G L	G R S A Q T E S P G	R F V L L D L A G E	A R T A G D						
GCCACCGCCG	GGGACGGCCT	GACGACCGGG	GACGCCACCG	TCGGCGGCAC	CTCTGGAGAC	GCCGCCCTCG	GCGACCGCC	CTCGGCTCGG	24400
A T A G D G L	T T G D A T V G G T S G D	A A L G S A L A T A L G S G							
GCGAGCCGCA	GCTCGCCCTC	CGGACGGGG	CGCTCCTCTGT	ACCCCGCCTG	GCGGGGGCCG	CCGCGCCCGC	CGGCGCGGAC	GGCCTGCGG	24500
E P Q L A L R D G A	L L V P R L A R A A A P A	A A D G L A A A D G							

FIG. 31WW

90/164

plkPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890
CCTCGCGCT	CTGCGGTGC	CCGCCGCTCC	GGCCCTCTGG	CGTCTGGAGC	CCGGTACGGA	CGGCAGCCTG	GAGAGCCTCA	CGGCGGCGCC	CGGCGACGCC
L A A	L P L P A A P	A A P A L W	R L E P G T D	G S L E S L T	A A P G D A				
24600									
GAGACCTCG	CCCCGAGCC	GCTCGGCCG	GGACAGGTCC	GCATCCGAT	CCGGGCCACC	GGTCTCACT	TCCGCGACGT	CCTGATCGCC	CTCGGCATGT
E T L A	P E P L G P G Q V R	I A I R A T	G L N F R D V	L I A L G M Y					
24700									
ACCCGATCC	GGCGTGATG	GGCACCGAGG	GAGCGGCGT	GGTCACCGG	ACCGGCCCG	GGTCACGCA	CCTCGCCCC	GGCGACCGG	TCATGGCCT
P D P A L M	G T E G A G V V T A	T G P G V T H	L A P G D R V	M G L					
24800									
GCTCTCGGC	GCGTACGCC	CGGTGTCGT	GGCGACGCG	CGGACGTCG	CGCGATGCC	CGAGGGTGG	ACGTTGCCC	AGGCGCCTC	CGTGCCGGTG
L S G A Y A P	V V V A D A R T V A	R M P E G W T F A Q	G A S V P V						
24900									
GTGTTCTGA	CGGCCGTCTA	CGCCCTGCG	GACCTGGCG	ACGTCAGCC	CGGCGAGCG	CTCCTGGTCC	ACTCCGCGC	CGGTGGCGTG	GGCATGGCCG
V F L T A V Y	A L R D L A D V K P	G E R L L V H S A A	G G V G M A A						
25000									

FIG. 31XX

91/164

plkPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890
CGGTGAGCT CGCCCGGCAC TGGGGCGTGG AGGTCCACGG CACGGCGAGT CACGGGAAGT GGGACGCCCT GCGCGCGCTC GGCCTGGACG ACGGCACAT	25100	V Q L A R H W G V E V H G T A S H G K W D A L R A L G L D D A H I							
CGCTCCTCC CGCACCTGG ACTTCGAGTC CGCGTTCGT GCCGTTCCG GCGGGCGGG CATGGACGTC GTACTGAAT CGCTCGCCCG CGAGTTCGTC	25200	A S S R T L D F E S A F R A A S G G A G M D V V L N S L A R E F V							
GACGCTCGC TCGGCTGCT CGGGCCGGGC GCGCGGTTG TGGAGATGG GAAGACCGAC GTCCGCGACG CGGAGCGGT CGCCGCGGAC CACCCCGGTG	25300	D A S L R L L G P G G R F V E M G K T D V R D A E R V A A D H P G V							
TCGGCTACCG CGCCTTCGAC CTGGGCGAGG CCGGGCCGGA GCGATCGC GAGATGCTG CCGAGTTCAT CGCCCTCTTC GAGGACGGG TGCTCCGCA	25400	G Y R A F D L G E A G P E R I G E M L A E V I A L F E D G V L R H							
CCTGCCCGTC ACGACTGGG ACGTGCGCG GCGCGCGGAC GCCTTCGGC ACGTCAGCA GCGCGCCAC ACGGGCAAG TCGTCTCAC GATGCCGTCG	25500	L P V T T W D V R R A R D A F R H V S Q A R H T G K V V L T M P S							

FIG. 31Y

92/164

plkPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890

PmlI
▼

GGCCTCGACC CGGAGGTAC GGTCTGTCTG ACCGGCGSCA CCGGTGCGCT GGGGGGCATC GTGGCCCGGC ACCTGTGGG CGAGTGGGC GTACGACGCC 25600
G L D P E G T V L L T G G T G A L G G I V A R H V V G E W G V R R L

ApaLI
▼

TGCTGCTCGT GAGCCGGCGG GGCACGGAGC CCCGGGGCGC CCGGAGCTC GTGCACGAGC TGGAGGCCCT GGGAGCCGAC GTCTCGGTGG CCGCGTGGCA 25700
L L V S R R G T D A P G A G E L V H E L E A L G A D V S V A A C D

CGTCGCCGAC CGGGAAGCCC TCACCGCGCT ACTCGACTCG ATCCCGCGCG AACACCCGCT CACCGGGTC GTCCACACGG CAGGCGTCCT CTCCGACGGC 25800
V A D R E A L T A V L D S I P A E H P L T A V V H T A G V L S D G

ACCTCCCTT CGATGACAGC GGAGGATGTG GAACACGTAC TGGCTCCCAA GGTCGACGCC GCGTTCCTCC TCGACGAACT CACCTCGACG CCGGGCTACG 25900
T L P S M T A E D V E H V L R P K V D A A F L L D E L T S T P G Y D

ACCTGGCAGC GTTCGTCATG TTCTCCTCCG CCGCCGCCGT CTTCTGTGGC GCGGGGCAGG GCGCCTACGC CGCCGCGCAAC GCCACCCCTCG ACGCCCTCGC 26000
L A A F V M F S S A A A V F G G A G Q G A Y A A A N A T L D A L A

FIG. 31ZZ

93/164

plkPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890

CTGCGCGCGC CGACAGCCG GACTCCCGC CCTCTCCCTC GGCTGGGGCC TCTGGCCGA GACCAGCGC ATGACCGCG GACTCAGCGA CACGACCGC 26100
 W R R R T A G L P A L S L G W G L W A E T S G M T G G L S D T D R

TGCGGCTGG CCCGTCGG GCGAGCGCC ATGACAGCG AGCTGACCCT GTCCCTCCTG GACGCGGCA TCGCGCGGA CGACCCGGCG CTCGTCCGA 26200
 S R L A R S G A T P M D S E L T L S L L D A A M R R D D P A L V P I

TGCGCCTGA CGTCGCGCG CTCGCGGCC AGCAGCGGA CGGCATGCTG GCGCGCTGC TCAGCGGCT CACCGCGGA TCGCGGTCG GCGGCGGCC 26300
 A L D V A A L R A Q Q R D G M L A P L L S G L T R G S R V G G A P

EspI SphI Bpu102I AscI
 ▼ ▼ ▼ ▼

GGTCAACGAG CGCAGGGCAG CCGCGCGAG CGCGGGCGAG ACCTCGCGG GCGGTCGCC GCGATGACAC CGGACGACCG GGTCGCGCAC 26400
 V N Q R R A A A G G A G E A D T D L G G R L A A M T P D D R V A H

PmlI
 ▼

CTGCGGACC TCGTCGTAC GCACGTGGG ACCGTCTCG GACACGGCAC CCCGAGCGG GTGACCTGG AGCGGCGCTT CCGGACACC GGTTCGACT 26500
 L R D L V R T H V A T V L G H G T P S R V D L E R A F R D T G F D S

FIG. 31AAA

94/164

plkPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890
CGCTCACCGC	CGTCGAATC	CGCAACGTC	TCAAGCCGC	GACCGGCTG	CGGCTGCCG	CCACGCTGGT	CTTCGACCAC	CCCACCCCGG	GGGAGCTCGC
L T A	V E L	R N R L	N A A	T G L	R L P A	T L V	F D H	P T P G	E L A
CGGGCACCTG	CTCGACGAAC	TCGCCACGGC	CGCGGGCGGG	TCCTGGGCGG	AAGGCACCGG	GTCCGGGAGAC	ACGGCTCGG	CGACCGATCG	GCAGACCACG
G H L	L D E L	A T A	A G G	S W A	E G T	G S G D	T A S A	T D R	Q T T
GGGGCCCTCG	CCGAACTCGA	CCGGCTGGAA	GGGTGCTCG	CCTCCCTCGC	GCCCGCCGCC	GGCGGCCGTC	CGGAGCTCGC	CGCCCGGCTC	AGGGCGCTGG
A A L	A E L D	R L E	G V L	A S L	A P A	A G G R P	E L A	A R L	R A L A
BstXI ▼									
CCGCGGCCCT	GGGGGACGAC	GGCGACGACG	CCACCGACCT	GGACGAGGGG	TCCGACGACG	ACCTCTTCTC	CTTCATCGAC	AAGGAGCTGG	GGGACTCCGA
A A L	G D D	G D D	A T D	L D E	A S D	D D L	F S F	I D K	E L G D S D
CTTCTGACCT	GCCCGACACC	ACCGGCACCA	CCGGCACCCAC	CAGCCCCCCT	CACACACGGA	ACACGGAACG	GACAGGCGAG	ACCGGGAGCC	ATGGCGAACA
F								M A N	N

FIG. 31BBB

95/164

plkPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890

PstI

SfcI

▼ ▼

ACGAAGACAA GCTCCGGGAC TACCTCAAGC GCGTCACCGC CGAGCTGCAG CAGAACACCA GCGTCTGCG CGAGATCGAG GGACGCACGC ACGAGCCGT 27100
 E D K L R D Y L K R V T A E L Q Q N T R R L R E I E G R T H E P V

GGCGATCGTG GGCATGGCCT GCCGCTGCC GGGCGGTGTC GCCTCGCCG AGGACCTGTG GCAGCTGGTG GCCGGGACG GGGACCGGAT CTCGGAGTTC 27200
 A I V G M A C R L P G G V A S P E D L W Q L V A G D G D A I S E F

MluI

▼

CCGCAGGACC GCGGCTGGGA CGTGGAGGGG CTGTACGACC CCGACCCGGA CCGTCCGGC AGGACGTACT GCCGTCGGG CGGATTCCTG CACGACGCCG 27300
 P Q D R G W D V E G L Y D P D P D A S G R T Y C R S G G F L H D A G

GCGAGTTCGA CGCCGACTTC TTCCGGATCT CGCCGCGCGA GGCCCTCGCC ATGGACCCGC AGCAGCGACT GTCCCTCACC ACCGCGTGG AGCGATCGA 27400
 E F D A D F F G I S P R E A L A M D P Q Q R L S L T T A W E A I E

GAGCGGGGC ATCGACCCGA CGGCCCTGAA GGGCAGCGGC CTCGGGTCT TCGTCGGCG CTGGCACACC GGCTACACT CGGGGCAGAC CACCGCCGTG 27500
 S A G I D P T A L K G S G L G V F V G G W H T G Y T S G Q T T A V

FIG. 31CCC

96/164

plkPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890
CAGTCGCCCG	AGCTGGAGGG	CCACCTGGTC	AGCGGCGCGG	CGCTGGGCTT	CCTGTCCGGC	CGTATCGCGT	ACGTCTCTCG	TACGGACGGA	CCGGCCCTGA
Q S P E L E G	H L V S G A A	L G F L S G R	I A Y V L G T	D G P A L T					
27600									
CGTGGACAC	GGCTGTCTCG	TCCTCGCTGG	TGCGCTGCA	CCTCGCGTG	CAGGCCCTCC	GCAAGGGCGA	GTGGACATG	GCCCTCGCG	GTGGTGTAC
V D T A C S S	S S L V A L H	L A V Q A L R	K G E C D M	A L A G G V T					
27700									
GGTCATGCCC	AACGGGACC	TGTTGTGCA	GTTGAGCCGG	CAGCGCGGC	TGGCCGCGGA	CGGCCGCTCG	AAGCGTTTCG	CCACCTCGGC	GGACGGCTTC
V M P N A D L	F V Q F S R Q	R G L A A D	G R S K A F A	T S A D G F					
27800									
GGCCCCCGGG	AGGCGCCGG	AGTCCTGCTG	GTGGAGCGCC	TGTCGAGCGC	CCGCCGCAAC	GGACACCGGA	TCCTCGCGGT	CGTCCGCGGC	AGCGCGTCA
G P A E . G A	G V L L V E R L	S D A R R N G	H R I L A V	V R G S A V N					
27900									
ACCAGGACGG	CGCCAGCAAC	GGCCTCACGG	CTCCGCACGG	GCCCTCCAG	CAGCGCGTCA	TCCGACGGGC	CCTGGGGAC	GCCCGGCTCG	CGCCGGGTGA
Q D G A S N G	L T A P H G P	S Q Q R V I	R R A L A D	A R L A P G D					
28000									

XmnI



BamHI



FIG. 31DDD

97/164

plkPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890
CGTGGACGTC	GTCGAGGCGC	ACGGCACGGG	CACGGGCTC	GGCGACCGA	TCGAGGCGCA	GGCCCTCATC	GCCACCTACG	GCCAGGAGAA	GAGCAGCGAA
V D V	V E A H	G T G	T R L	G D P I	E A Q	A L I	A T Y G	Q E K	S S E
CAGCCGCTGA	GGCTGGGCGC	GTTGAAGTCG	AACATCGGGC	ACACGCAGGC	CGCGGCCCGT	GTCGCAAGTG	TCATCAAGAT	GGTCCAGGCG	ATGCGCCACG
Q P L R	L G A	L K S	N I G H	T Q A	A A G	V A G V	I K M	V Q A	M R H G
GACTGCTGCC	GAAGACGCTG	CACGTGCGACG	AGCCCTCGGA	CCAGATCGAC	TGGTCGGCGG	GCACGGTGA	ACTCCTCACC	GAGGCCGTCG	ACTGGCCGGA
L L P	K T L	H V D	E P S D	Q I D	W S A G	T V E	L L T	E A V D	W P E
GAAGCAGGAC	GGCGGGCTGC	GCCGCGCGGC	TGTCTCTCC	TTCCGCATCA	GCGGACGAA	CGCGCACGTC	GTCCTGGAGG	AGGCCCCCGC	GGTCGAGGAC
K Q D	G G L R	R A A	V S S	F G I S	G T N	A H V	V L E E	A P A	V E D
TCCCCGGCCG	TCGAGCCGCC	GGCCGGTGGC	GGTGTGGTGC	CGTGGCCGCT	GTCGCGAAG	ACTCCGGCCG	CGCTGGACGC	CCAGATCGGG	CAGCTCGCCG
S P A V	E P P	A G G	G V V	P W P V	S A K	T P A	A L D A	Q I G	Q L A A

FIG. 31EEE

98/164

plkPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890

BamHI
▼

CGTACGCGGA CGGTCGTACG GACGTGGATC CGGCGGTGGC CGCCCGCGCC CTGGTCGACA GCCGTACGGC GATGGAGCAC CGGCGGTGGC CGGTCGCGCA 28600
 Y A D G R T D V D P A V A A R A A R A L V D S R T A M E H R A V A V G D

CAGCCGGGAG GCACTGCGGG ACGCCCTGCG GATGCCGGA GGA CTGGTAC GCGCAGCTC CTCGGACGTG GGCCGGGTGG CGTTCGTCTT CCCCGGCCAG 28700
 S R E A L R D A L R M P E G L V R G T S S D V G R V A F V F P G Q

BsmI
▼

GGCAGCAGT GGGCCGGCAT GGGCGCCGAA CTCCTTGACA GCTCACCAGA GTTCGCTGCC TCGATGGCCG AATGGAGAC CGCGCTCTCC CGCTACGTGG 28800
 G T Q W A G M G A E L L D S S P E F A A S M A E C E T A L S R Y V D

ACTGGTCTCT TGAAGCCGTC GTCCGACAGG AACCCGGCGC ACCCAGCTC GACCGCGTCG ACGTCGTCCA GCCCGTGACC TTGCTGTCA TGGTCTCGCT 28900
 W S L E A V V R Q E P G A P T L D R V D V V Q P V T F A V M V S L

ApaLI
▼

GGCGAAGGTC TGGCAGCACC ACGGCATCAC CCCCAGGCC GTGTCGGCC ACTCGCAGG CGAGATCGCC GCCCGGTGCG TCGCGGTGC ACTCACCTC 29000
 A K V W Q H H G I T P Q A V V G H S Q G E I A A A Y V A G A L T L

FIG. 31FFF

99/164

plkPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890
GACGACGCCG	CCC	CGT	CGT	CGT	CGT	CGT	CGT	CGT	CGT
D D A A	R V V	T L R	S K S	I A A	H L A	G K G	M I S	L A L	D E A
29100									
TCCTGAAGCG	ACT	GAG	CGAC	TT	CGAC	GGAC	TCT	CGT	CGC
L K R L	S D F	D G L	S V A	A V N	G P T	A T V	V S G	D P T	Q I E
29200									
CGCCCGCACC	TGC	GAG	GCGC	ACG	GCG	GTC	CG	ACG	CGC
A R T C	E A D	G V R	A R I	I P V	D Y A	S H S	R Q V	E I I	E K E
29300									
GAGTCTCTCG	CCG	GACT	CGC	CCC	GAG	GCT	CGC	GGT	GCT
E V L A	G L A	P Q A	P H V	P F F	S T L	E G T	W I T	E P V	L D G
29400									
GGTACCGCAA	CCT	GCG	CCAT	CGG	TGG	GGT	CGT	ACG	GCT
Y R N L	R H R	V G F	A P A	V E T	L A V	D G F	T H F	I E V	S A H
29500									

FIG. 31GGG

100/164

plkPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890
CCTCACCATG	ACCCTCCCG	AGACCGTCAC	CGGCTCCGC	ACCCTCCGC	GCGAACAGGG	AGGCCAGGAG	CGTCTGGTCA	CCTCACTCGC	CGAAGCCTGG
L T M	T L P E	T V T	G L G	T L R R	E Q G	G Q E	R L V T	S L A	E A W
39600									
GCCAACGGCC	TCACCATCGA	CTGGGCGGCC	ATCCTCCCA	CCGCAACCGG	CCACACCCC	GAGCTCCCCA	CCTACGCCTT	CCAGACCGAG	CGCTTCTGGC
A N G L	T I D	W A P	I L P T	A T G	H H P	E L P T	Y A F	Q T E	R F W L
29700									
TGCAGAGCTC	CGGCGCCACC	ACGCGCGCCG	ACGACTGGCG	TTACCGCGTC	GAGTGAAGC	CGCTGACGGC	CTCCGGCCAG	GCGGACCTGT	CCGGGCGGTG
Q S S	A P T	S A A	D W R	Y R V	E W K P	L T A	S G Q	A D L S	G R W
29800									
GATCGTCGCC	GTCCGGGAGCG	AGCCAGAAGC	CGAGCTGCTG	GGCGCGCTGA	AGGCCGCGGG	AGCGGAGGTC	GACCTACTGG	AAGCCGGGCG	GGACGACGAC
I V A	V G S E	P E A	E L L	G A L K	A A G	A E V	D V L E	A G A	D D D
29900									
CGTGAGGCCC	TCGCGCGCCG	GTCACCGCA	CTGACGACCG	GCGACGGCTT	CACCGGCTG	GTCTCGCTCC	TCGACGACCT	CGTGCCACAG	GTGCGCTGGG
R E A L	A A R	L T A	L T T G	D G F	T G V	V S L L	D D L	V P Q	V A W V
30000									

FIG. 31HHH

101/164

pikPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890

TGCAGGCACT CGGCGACGCC GGAATCAAGG CGCCCCCTGTG GTCCGTCACC CAGGGCGCGG TCTCGTGGG ACGTCTCGAC ACCCCCGCGG ACCCGGACCG 30100
Q A L G D A G I K A P L W S V T Q G A V S V G R L D T P A D P D R

GGCCATGCTC TGGGGCCTCG GCCCGCTCGT CGCCCTTGAG CACCCCGAAC GCTGGCGCGG CCTCGTCGAC CTCCCGCGCC AGCCCGATGC CGCGGCCCTC 30200
A M L W G L G R V V A L E H P E R W A G L V D L P A Q P D A A A L

BsaBI



GCCCACCTCG TCACGGCACT CTCCGGCGCC ACGGGCGAGG ACCAGATCGC CATCCGCACC ACCGGACTCC ACGCGCGCGG CCTCGCGCGG GCACCCCTCC 30300
A H L V T A L S G A T G E D Q I A I R T T G L H A R R L A R A P L H

ACGGACGTCG GCCCACCCTCG GACTGGCAGC CCCACGGCAC CGTCCTCATC ACCGGCGGCA CCGGAGCCCT CGGCAGCCAC GCGGCACGCT GGATGGCCCA 30400
G R R P T R D W Q P H G T V L I T G G T G A L G S H A A R W M A H

CCACGGAGCC GAACACCTCC TCCTCGTCAG CCGCAGCGGC GAACAAGCCC CCGGAGCCAC CCAACTCACC GCGGAACTCA CCGCATCGG CGCCCCGCGT 30500
H G A E H L L L V S R S G E Q A P G A T Q L T A E L T A S G A R V

FIG. 31III

[illegible]

ACCATGCGCCTGTGCGACGT CGCGGACCCC CACGCGCATGC GCACCCCTCCT CGACGCCATC CCGCGCAGA CGCCCTCAC CGCGTGTGTC CACACCGCCG 30600
T I A A C D V A D P H A M R T L L D A I P A E T P L T A V V H T A G

GCGCACCGG CGCGCATCCG CTGCACGTCA CCGGCCGGA GGACATCGCC CGCATCTGG GCGCGAGAC GAGCGCGCC GAGGTCTCG ACGACCTGCT 30700
A P G G D P L D V T G P E D I A R I L G A K T S G A E V L D D L L

CGCGGGCACT CCGCTGGACG CCTTCGTCTT CTACTCTCTG AACGCCGGG TCTGGGGCAG CGGAGGCAG GGGTCTACG CGCGGGCAA CGCCACCTC 30800
R G G T P L D A F V L Y S S N A G V W G S G S Q G V Y A A A N A H L

INTM

GACGCGCTCG CCGCCCGCGG CCGCGCCGG GCGGAGACGG CGACTCTCGT CGCTGGGGC CTCTGGGGC GCGACGGCAT GGGCTGGGGC GCCGACGACG 30900
D A L A A R R R A R G E T A T S V A W G L W A G D G M G R G A D D A

MacI

Bali

CGTACTGGCA GCGTCGCGGC ATCCGTCGGA TGAGCCCGG CCGGCGCCCTG GAGCAACTGG CCAAGGCCCT GAGCCACGAC GAGACCTTCG TCGCCGTGGC 31000
Y W Q R R G I R P M S P D R A L D E L A K A L S H D E T F V A V A

FIG. 31JJ

103/164

plkPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890

CGATGTCGAC TGGGAGCGGT TCGCGCCCGC GTTCACGGTG TCCCGTCCCA GCCTTCTGCT CGACGGCGTC CCGGAGGCC GGCAGGCGT CGCCGCACCC 31100
 D V D W E R F A P A F T V S R P S L L L D G V P E A R Q A L A A P

GTCGGTGCCC CGGCTCCCGG CGACGCCCGC GTGGGCCGGA CCGGGCAGTC GTGGCGCTG GCCGCGATCA CCGCGCTCCC CGAGCCGAG CGCCGGCCGG 31200
 V G A P A P G D A A V A P T G Q S S A L A A I T A L P E P E R R P A

CGCTCCTCAC CCTCGTCGGT ACCACGCGG CGGCGTACT CGGCCATTCC TCCCGCGACC GGGTGGCCC CGGCGTGCC TTCACGAGC TCGGCTTCGA 31300
 L L T L V R T H A A A V L G H S S P D R V A P G R A F T E L G F D

CTCGCTGACG GCCGTGCAGC TCCGCAACCA GCTCTCCAG GTGGTCGGCA ACAGGCTCCC CGCCACCAG GTCTTCGACC ACCGACGCC CGCCGCACCTC 31400
 S L T A V Q L R N Q L S T V V G N R L P A T T V F D H P T P A A L

GCCGCGCACC TCCACGAGGC GTACCTCGCA CCGGCCGAGC CGGCCCCGAC GGACTGGGAG GGGCGGGTGC GCCGGGCCCT GGCCGAACTG CCCCTCGACC 31500
 A A H L H E A Y L A P A E P A P T D W E G R V R R A L A E L P L D R

FIG. 31KKK

104/164

plkPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890
GGCTGCGGGA	CGCGGGGTC	CTCGACACCG	TCCTGCGCCT	CACCGGCATC	GAGCCCGAGC	CGGTTCCGG	CGGTTCCGAC	GGCGGCGCCG	CCGACCCCTGG
L R D	A G V	L D T V	L R L	T G I	E P E	P G S G	G S D	G G A	A D P G
TGCGGAGCCG	GAGGGTTCGA	TCGACGACCT	GGACGCCGAG	GCCCTGATCC	GGATGGCTCT	CGGCCCCCGT	AACACCTGAC	CCGACCGCGG	TCCTGCCCCA
A E P	E A S I	D D L	D A E	A L I	R M A	L G P	R N T		
CGCGCCGCAC	CCCGGCGATC	CGCGGCACCA	CCCGCCCCCA	CAGCCCCACA	ACCCCATCCA	CGAGCGGAAG	ACCACACCCA	GATGACGAGT	TCCAACGAAC
								M T S	S N E Q
ACTTGGTGGA	CGCTCTGCGC	GCCTCTCTCA	AGGAGAACGA	AGAACTCCGG	AAAGAGAGCC	GTGCGCGGCG	CGACCGTCGG	CAGGAGCCCA	TGGCGATCGT
L V D	A L R	A S L	K E N	E E L	R K E	S R R	R A D	R R Q	E P M A I V
CGGCATGAGC	TGCCGGTTCG	CGGGCGGAAT	CCGTCCTCCC	GAGGACCTCT	GGGACGCCGT	CGCCGCGGCG	AAGGACCTGG	TCTCCGAGGT	ACCGGAGGAG
G M S	C R F A	G G I	R S P	E D L	W D A	V A A	G K D	L V S	E V P E E

KpnI
Acc65I

FIG. 31LLL

105/164

plkPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890
CGCGGCTGGG	ACATCGACTC	CCTCTACGAC	CCGCTGCCCC	GGCGCAAGGG	CACGACGTAC	GTCGCAACG	CCGCTTCCT	CGACGACGCC	GCCGGATTCG
R G W D	I D S	L Y D	P V P G	R K G	T T Y	V R N A	A F L	D D A	A G F D
ACGCGGCCTT	CTTCGGGATC	TCGCCGCGCG	AGGCCCTCGC	CATGGACCCG	CAGCAGCGGC	AGCTCCTCGA	AGCCTCCTGG	GAGGTCTTCG	AGCGGCGCCG
A A F	F G I	S P R E	A L A	M D P	Q Q R Q	L L E	A S W	E V F	E R A G
CATCGACCCC	CGTCGGGTCC	GCGGCACCGA	CGTCGGCGTG	TACGTGGGCT	GTGCTACCA	GGACTACGG	CCGGACATCC	GGGTGCCCC	CGAAGGCACC
I D P	A S V R	G T D	V G V	Y V G C	G Y Q	D Y A	P D I R	V A P	E G T
GGCGGTACG	TCGTCACCGG	CAACTCCTCC	GCCGTGGCCT	CCGGCGCAT	CGCGTACTCC	CTCGGCTGG	AGGGACCCGC	CGTGACCGTG	GACACGGCGT
G G Y V	V T G	N S S	A V A S	G R I	A Y S	L G L E	G P A	V T V	D T A C
GCTCCTCTTC	GCTCGTCGCC	CTGCACCTCG	CCCTGAAGGG	CCTGCGGAAC	GGCGACTGCT	CGACGGCACT	CGTGGCGCGC	GTGGCCGTCC	TCGCGACGCC
S S S	L V A	L H L A	L K G	L R N	G D C S	T A L	V G G	V A V	L A T P

FIG. 31MMM

106/164

plPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890

GGGCGGTTTC ATCGAGTTCA GCAGCCAGCA GGCCATGGCC GCCGACGGCC GGACCAAGGG CTTCGCTCG GCGGCGGACG GCCTCGCCTG GGGCGAGGGC 32600
G A F I E F S S Q Q A M A A D G R T K G F A S A A D G L A W G E G

GTGCGCGTAC TCCTCCTCGA ACGGCTCTCC GACGGCGGC GCAAGGGCCA CCGGCTCCTG GCCGCTGTC GCGGCGGCGC CATCAACCAG GACGGCGCGA 32700
V A V L L L E R L S D A R R K G H R V L A V V R G S A I N Q D G A S

GCAACGGCCT CACGGCTCCG CACGGGCCCT CCGAGCAGCA CCTGATCCGC CAGGCCCTGG CCGACGGCG GCTCAGCTCG AGCGACGTGG ACGTCGTGGA 32800
N G L T A P H G P S Q Q H L I R Q A L A D A R L T S S D V D V V E

Asci
▼

GGGCCACGGC ACGGGACCC GTCTCGGCGA CCGGATCGAG GCGCAGGCGC TGCTGCCAC GTACGGCGAG GGGCGGCGCC CCGGCGGCGC GCTGCGGCTG 32900
G H G T G T R L G D P I E A Q A L L A T Y G Q G R A P G Q P L R L

GGGACGCTGA AGTCGAACAT CCGGCACACG CAGGCCGCTT CCGGTGTGCG CGGTGTATC AAGATGGTGC AGGCGCTGCG CCACGGGGTG CTGCCGAAGA 33000
G T L K S N I G H T Q A A S G V A G V I K M V Q A L R H G V L P K T

FIG. 31NNN

107/164

plkPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890

PmlI

▼ CCCTGCACGT GGACGAGCCG ACGGACCAGG TCGACTGGTC GGGCGGTTTCG GTCGAGCTGC TCACCGAGGC CGTGGACTGG CCGGAGCGGC CCGGCCCGGCT 33100
 L H V D E P T D Q V D W S A G S A G S V E L L T E A V D W P E R P G R L

CCGCCGGGCG GCGCTCTCCG CGTTCGGCGT GGGCGGGACG AACGGGCACG TCGTCTGGA GGAGGCCCCG GCGGTCGAGG AGTCCCTGTC CGTCGAGCCG 33200
 R R A G V S A F G V G G T N A H V V L E E A P A V E E S P A V E P

CCGGCCCGGTG GCGGCTGGT GCGTGGCCG GTGTCCGCGA AGACTCGGC CGCACTGGAC GCCCAGATCG GGCAGCTGTC CGCATACGCG GAAGACCGCA 33300
 P A G G G V V P W P V S A K T S A A L D A Q I G Q L A A Y A E D R T

BamHI

▼ CGGACGTGGA TCCGGCGGTG GCCGCCCGCG CCTTGGTCTGA CAGCCGTACG GCGATGGAGC ACCGCGCGGT CCGGTCGTC GACAGCCGGG AGGCACTGCG 33400
 D V D P A V A A R A L V D S R T A M E H R A V A V G D S R E A L R

GGACGCCCTG CGGATGCCG AAGGACTGGT ACGGGGCACG GTCACCGATC CCGGCCGGGT GCGGTTCTGTC TTCCCGCGCC AGGGCACGCA GTGGGCCCGC 33500
 D A L R M P E G L V R G T V T D P G R V A F V F P G Q G T Q W A G

FIG. 31000

108/164

pikPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890

EcoRI
ApoI

BsmI

ATGGGCGCG AACTCCTCGA CAGCTCACCC GAATTGCGCG CGGCCATGGC CGAATGCGAG ACCGCACTCT CCCCGTACGT CGACTGGTCT CTCGAAGCCG 33600
M G A E L L D S S P E F A A A M A E C E T A L S P Y V D W S L E A V

TCGTCCGACA GGCTCCCGAGC GCACCGACAC TCGACCGCGT CGACGTCGTC CAGCCCGTCA CCTTCGCGGT CATGGTCTCC CTGCGCAAGG TCTGGCAGCA 33700
V R Q A P S A P T L D R V D V V Q P V T F A V M V S L A K V W Q H

CCACGGCATC ACCCCCGAGG CCGTCATCGG CCACTCCGAG GCGAGATCG CCGCCGCGTA CGCTCACCC TCGACGACGC CGCTCGTGTC 33800
H G I T P E A V I G H S Q H S Q G E I A A A Y V A G A L T L D D A A R V

GTGACCTCC GCAGCAAGTC CATGCGCGCC CACCTCGCCG GCAAGGCGG CATGATCTCC CTCGCCCTCA GCGAGGAAGC CACCCGGCAG CGCATCGAGA 33900
V T L R S K S I A A H L A G K G G M I S L A L S E E A T R Q R I E N

ACCTCCACGG ACTGTGATC GCGCGCGTCA ACGGGCCTAC CGCCACCGTG GTTTCGGGCG ACCCCACCCA GATCCAAGAA CTGTCTCAGG CGTGTGAGGC 34000
L H G L S I A A V N G P T A T V V S G D P T Q I Q E L A Q A C E A

FIG. 31PPP

109/164

plkPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890

CGACGGCATC CGCGCACGGA TCATCCCCGT CGACTACGCC TCCACAGCG CCCACGTGGA GACCATCGAG AACGAACCTG CCGACGTCCT GCGGGGTTG 34100
 D G I R A R I I P V D Y A S H S A H V E T I E N E L A D V L A G L

KpnI
 Acc65I
 ▼

TCCCCCAGA CACCCAGGT CCCCTTCTTC TCCACCTCG AAGCACCTG GATCACGGA CCCGCCCTCG ACGGCGGCTA CTGGTACCGC AACCTCGCC 34200
 S P Q T P Q V P F F S T L E G T W I T E P A L D G G Y W Y R N L R H

ATCGTGTGGG CTTGCGCCCG GCGTTCGAGA CCTCGGCCAC CGACGAAGGC TTCACCCACT TCATCGAGGT CAGCGCCAC CCGTCTCTCA CCATGACCT 34300
 R V G F A P A V E T L A T D E G F T H F I E V S A H P V L T M T L

MscI
 BalI
 ▼

CCCCGACAAAG GTACCGGCC TGGCCACCCT CCGACGCGAG GACGGGGAC AGCACCGCT CACCACCTCC CTGCGGAGG CTTGGGCCAA CGGCCTCGCC 34400
 P D K V T G L A T L R R E D G G Q H R L T T S L A E A W A N G L A

CTCGACTGGG CCTCCCTCCT GCCCGCCACG GCGGCCCTCA GCCCGCCGT CCGGACCTC CCGACGTAG CCTTCCAGCA CCGCTCGTAC TGGATCAGCC 34500
 L D W A S L L P A T G A L S P A V P D L P T Y A F Q H R S Y W I S P

FIG. 31QQQ

110/164

plkPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890
CGCGGGTCC	CGGCGAGGG	CGCGGCACA	CGGTTCCGG	GCGCGAGGC	GTCGCCGGA	CGGGGCTCG	GTGGGGCCG	GGTGCCGAG	ACCTCGACGA
A G P G E A	P A H T A S G	R E A V A E T	G L A W G P	G A E D L D E					
GGAGGGCCG	CGCAGCGCG	TACTCGCGAT	GGTGATGCG	CAGCGGCCT	CCGTGCTCG	GTGCGACTCG	CCCGAAGAG	TCCCCGTCA	CGCCCCGCTG
E G R R S A	V L A M V M R	Q A A S V L R	C D S P E E V	P V D R P L					
CGGGAGATCG	GCTTCGACTC	GCTGACCGCC	GTCGACTTCC	GCAACCGCGT	CAACCGGCTG	ACCGGTCTCC	AGTGCCTCC	CACCGTCGTG	TTCCAGCAC
R E I G F D S	L T A V D F R	N R V N R L	T G L Q L P P	T V V F Q H P					
	*								ACP 6
CGACGCCCGT	CGGCTCGCC	GAGCGCATCA	GCGACGAGCT	GGCCGAGCGG	AACGCGGCG	TCGCCGAGCC	GTCGGATCAC	GAGCAGGCGG	AGGAGGAGAA
T P V A L A	E R I S D E L	A E R N W A V	A E P S D H	E Q A E E K					TE domain
GGCCGCCCGT	CGGCGGGG	CCCGTCCGG	GGCCGACACC	GGCGCCGCG	CCGGATGTT	CCGCCCTG	TTCCGGCAGG	CCGTGGACGA	CGACCGGTAC
A A A P A G A	R S G A D T	G A G A G M F	R A L F R Q A	V E D D R Y					

FIG. 31RRR

111/164

plkPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890

GGCGAGTTCC TCGACGTCCT CGCCGAAGCC TCCGGGTTC TCCGCTGCCC GAGGCCTGCT CGGAGCGGCT CGACCCGGTG CTGCTGCGCG 35100
G E F L D V L A E A S A F R P Q F A S P E A C S E R L D P V L L A G

EspI
Bpu1102I
▼

GGGTCCGAC GGACCGGGCG GAAGGCCGTG CCGTTCTCGT CGGCTGCACC GGCACCGCGG CGAACGGCGG CCGGACGAG TTCCTGCGGC TCAGCACCTC 35200
G P T D R A E G R A V L V G C T G T A A N G G P H E F L R L S T S

CTTCCAGGAG GAGCGGACT TCCTCGCCGT ACCTCTCCCC GGCTACGGCA CGGTACGGG CACCGGCACG GCCCTCCTCC CGGCCGATCT CGACACCGCG 35300
F Q E E R D F L A V P L P G Y G T G T G T G T A L L P A D L D T A

CTCGACGCC AGGCCCGGC GATCTCCGG GCCCGCGGG ACGCCCGGT CGTCTGCTC GGGCGGCCCT GCTCGCGCAC GAGCTGGCCT 35400
L D A Q A R A I L R A A G D A P V V L L G H S G G A L L A H E L A F

AscI
▼

TCCGCTTGA GCGGGCGCAC GCGCGCGGAT CGTCTGCTC GACCCCTATC CGCCGGGCCA TCAGGAGCCC ATCGAGGTGT GGAGCAGGCA 35500
R L E R A H G A P P A G I V L V D P Y P P G H Q E P I E V W S R Q

FIG. 31SSS

112/164

plkPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890

MscI
 Ball

GCTGGGCGAG GGCCTGTTCG CGGGCGAGCT GGAGCGGATG TCCGATGCGC GGCTGCTGGC CATGGGCGCG TACGGCGGGT TCCTGCGCGG CCGCGGCGCG 35600
 L G E G L F A G E L E P M S D A R L L A M G R Y A R F L A G P R P

GGCGGCGAGCA GCGCGCCCGT GCTTCTGGTC CGTGGCTCCG AACCGCTGGG CGACTGGCAG GAGGAGCGGG GCGACTGGCG TGCCCACTGG GACCTTCCGC 35700
 G R S S A P V L L V R A S E P L G D W Q E E R G D W R A H W D L P H

ACACCGTCGC GGACGTGCG GCGGACCACT TCACGATGAT GCGGACCACT GCGCGGCGCG TCGCGGAGGC CGTCTCTCC TGGCTCGACG CCATCGAGGG 35800
 T V A D V P G D H F T M M R D H A P A V A E A V L S W L D A I E G

BamHI

CATCGAGGGG GCGGGCAAGT GACCGACAGA CCTCTGAACG TGGACAGCG ACTGTGGATC CGGGGCTTCC ACCCGCGCC GAACAGCGCG GTGCGGCTGG 35900
 I E G A G K V T D R P L N V D S G L W I R R F H P A P N S A V R L V
 →

TE2

TCTGCCTGCC GCACGCCGCG GGCTCCGCCA GCTACTTCTT CCGCTTCTCG GAGGAGCTGC ACCCTCTCCGT CGAGGCCCTG TCGGTGCACT ATCCGGGCGG 36000
 C L P H A G G S A S Y F F R F S E E L H P S V E A L S V Q Y P G R

FIG. 31TTT

113/164

plPKPS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890

CCAGGACCGG CGTGCCGAGC CGTGCTGGA GAGCTGCGG GAGCTGTGGT CGCGGCCACC GAACCTGGT GGCAGGAGGG CCGGCTGGCC 36100

Q D R R A E P C L E S V E E L A E H V V A A T E P W W Q E G R L A

TTCTTCGGC ACAGCTCGG CGCTCCGTC GCCTTCGAGA CGGCCCGCAT CCTGGAACAG CGGCACGGG TACGGCCCGA GGGCCTGTAC GTCTCCGTC 36200

P F G H S L G A S V A F E T A R I L E Q R H G V R P E G L Y V S G R

▼
 AscI
 EspI
 Bpu1102I

GGCGGCCCC GTGCTGGCG CCGGACCGG TCGTCCACCA GCTGGAGAC CGGCGGTTC TGGCCGAGAT CCGGCGGCTC AGCGGCACCG ACGAGCGGTT 36300

R A P S L A P D R L V H Q L D D R A F L A E I R R L S G T D E R F

CCTCCAGGAC GACGAGCTGC TGGCGTGGT GCTGCCCGG CTGCGCAGCG ACTACAAGG GCGGAGAGC TACCTGCACC GCGCGTCCG CAAGCTCACC 36400

L Q D D E L L R L V L P A L R S D Y K A A E T Y L H R P S A K L T

TGCCCCGTGA TGCCCTGGC CGGCGACCGT GACCGGAAG CGCGGTGAA CGAGTGGCC GAGTGGGTC GGCACACCAG CCGGCGGTC TGCCTCCGG 36500

C P V M A L A G D R D P K A P L N E V A E W R R H T S G P F C L R A

FIG. 31UUU

114/164

plkPKS Sequence

10	20	30	40	50	60	70	80	90	100
1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890
CGTACTCCGG CGGCCACTTC TACCTCAACG ACCAGTGGCA CGAGATCTGC AACGACATCT CCGACCACTT GCTGTCACC CGCGGCGGC CCGATGCCCG 36600									
Y S G G H F Y L N D Q W H E I C N D I S D H L L V T R G A P D A R									
CGTCGTGCAG CCCCAGACCA GCCTTATCGA AGGAGCGGCG AGAAGATGGC AGAAGCCACG GTGACCGAGC ACCTGACGGG GGCCCTCAGC CAGCCCCCGC 36700									
V V Q P P T S L I E G A A K R W Q N P R									
TGGGCGGCAC CGTCGGCGCG GTGGCGGACC GTGAATCTCG CACCCACCTC CTGGAGAGCC GCGGCATCCA CTGGATCC 36778									

FIG. 31W

115/164

sugar.finalgene b-1 Sequence

10 20 30 40 50 60 70 80 90
123456789012345678901234567890123456789012345678901234567890

BamHI

▼
GGATCCGGCGCTTCCACCCCGCGCCGAACAGCGCGGTGCGGCTGTCTGCCCTGCCGACGCCGGGGGCTCCGCCAGCTACTTCTCCGCT
CCTAGGCCGCGAAGGTGGGGCGCGGCTTGTTCGGGCCACGCCGACCCAGACCGGCGGTGCGGCCCGCCGAGCGGTTCGATGAAGAGCGGA
I R R F H P A P N S A V R L V C L P H A G S A S Y F F R F

90

TCCTCGGAGGAGTGACCCCTCCGTCGAGGCCCTGTGCGTGCAATCCGGGCCGCCAGGACCGGCGTCCCGAGCCCGTGTCTGGAGAGCG
AGACCTCCTCGACGTGGGGAGGCAGCTCCGGGACAGCCACGTATAGGCCCGCGGTCTCTGGCCGACCGGTTCGGCACAGACCTCTCGC
S E E L H P S V E A L S V Q Y P G R Q D R R A E P C L E S V

180

NspHI

▼
TCGAGGAGCTCGCCGAGCATGTGGTCGGGCCACCGAACCCCTGTGGCAGGAGGCCCGGCTGGCCCTTCTTCGGGCACAGCCTCGGGCGCT
AGCTCCTCGAGCGGCTCGTACACAGCGCCGGTGGCTTGGGACCAACCGTCTCCCGGCCGACCCGGAAGAACCCCGTGTTCGGAGCCCGGA
E E L A E H V V A A T E P W Q E G R L A F F G H S L G A S

270

AscI

▼
CCGTCCGCTTCGAGACGGCCCGCATCCTGGAAACAGCGGCACGGGGTACGGCCCGAGGGCCTGTACGTCTCCGGTCCGGCGGCCCGCTCGC
GGCAGCGGAAGCTCTGCCGGGCGTAGACCTTGTTCGCCGTGCCCATGCCGGCTCCCGGACATGCAGAGGCCAGCCCGCGGGGCGAGCG
V A F E T A R I L E Q R R H G V R P E G L Y V S G R R A P S L

360


FIG. 32A

116/164

[illegible]

RsrII	▼	PvuII	EspI	
		PflMI	Bpu1102I	
		▼	▼	
				450

TGGCGCCGGACCGGCTCGTCCACCAAGCTGGACGACCGGGCGTTCTTGGCCGAGATCCGGGGGCTCAGCGGCACCGACGAGCGGTTCCTCC
 ACCGCGGCTGGCCGACGAGGTGTCGACCTGCTGGCCCCGCAAGGACCGGCTCTAGGCGGCCGAGTCGCCGTGGCTGCTCGCCAAGGAGG
 A P D R L V H Q L D D R A F L A E I R R L S G T D E R F L Q

FspI  540

AGGACGACGAGTGTGTCGGGTGTCGCCGGCTCGCGACGCACTACAAGCGCGGAGACGTACCTGCACCGGCCGTCGCCCAAGC
TCCTGTGTCGACGACCGCGACCAACGACGGGGCGGACGGGTGCTGATGTTCCGCCGCTCTGCATGGACGTGGCCGCGCAGCGGTTCCG
D D E L L R L V L P A L R S D Y K A A E T Y L H R P S A K L

630

BglII
▼
CGTTCTGCTCCGGGCGTACTCCGGGGCCACTTCTACCTCAACGACCAGTGGCAACGAGATCTGCAACGACATCTCCGACCACCTGCTCG
GCAAGACGGAGGCCCGCATGAGGCCCGCGGTGAAGATGGAGTTGCTGGTCAACCGTGTCTAGACGTTGCTGTAGAGGCTGGTGGACGAGC
F C L R A Y S G G H F Y L N D Q W H E I C N D I S D H L L V

FIG. 32B

117/164

sugar.finalgene b-1 Sequence

10	20	30	40	50	60	70	80	90
1234567890123456789012345678901234567890123456789012345678901234567890								
AscI ▼ TCACCCGGCGCGCCGATGCCCGCGTCGTGACGCCCGACACGCTTATCGAAGGAGCGGCGAAGAGATGGCAGAAACCCACGGTGAC AGTGGCGCGCGCGGCTACGGGCGCAGACGTCGGGGCTGGTCGGAATAGCTTCCCTCGCCGCTTCTCTACCGTCTTGGGTGCCACTG T R G A P D A R V V Q P P T S L I E G A A K R W Q N P R V T								
EarI ▼ BstEII ▼								
810								
Apal ▼ CGACGACCTGACGGGGCCCTCACGAGCCCGCTGGCGCGCACCGTCCGCGGTGGCCGACCGTGAACTCGGCACCCACCTCCTGGA GCTGCTGGACTGCCCGCGGAGTCGTGCGGGGCGACCCGGCTGGCAGCGCGCCACCGGCTGGCACTTGAGCCGCTGGGTGGAGGACCT								
PflMI ▼								
900								
BamHI ▼ D D L T G A L T Q P P L G R T V R A V A D R E L G T H L L E								
GACCCGGGCATCCACTGGATCCACGCCCGGAACGGCGACCCGTACGCCACCGTGTGCGGCGCAGCGGACGACCCGTATCCCGCGTA CTGGCGCGGTAGGTGACCTAGGTGCGGGCGCTGCCGTGGGCATGCGGTGGCACGACCGCGCGGTCCGCCCTGCTGGGCATAGGGCGCAT								
990								
T R G I H W I H A A N G D P Y A T V L R G Q A D D P Y P A Y								
BstEII PvuII ▼								
1080								
CGAGCGGGTGCCTCCCGCGCGGCTCTCCTTACGCCCGACGGGACGCTGGGTCAACGCCGATCAAGCCCTGGCGCGGAGCATCCTCTG GCTCGCCACGACGGGCGCGCGGAGAGGAAGTCGGGCTGCCGTGACCCAGTGGCGGCTAGTGGGGACCGCCGCTCGTAGGAGAC E R V R A R G A L S F S P T G S W V T A D H A L A S I L C								

FIG. 32C

118/164

[illegible]

S S T D F G V S G A D G V P V P Q Q V L S Y G E G C P L E R E E

GCAGGTGCTCCGGCGGCCGGTGACGTGCCCGGAGGGCGGGCAGCGTGCTCGAGGGATCCACCGGGAGACGCTGGAGGGTCTCGC
CGTCCACGACGGCCCGCGGCCACTGCACGGCCCTCCCGCCCGTTCGACGGCACCAAGCTCCCTTAGGTGGCCCTCTCGGACCTCCQAGAGCG

Q V L P A A G D V P E G G Q R A V V E G I H R E T L E G L A

GCCGGAACCGTCCGGCTCGTACGCCCTTCGAGCTGCTGGGGCGTTTCGTCCGCCCGGGCGGTACCGGCCGCTGCCGCCCGCGTCTGGGTGT
CGGCCTGGGCAGCCGCAGCATGCGGAAGCTCGACGACCCGCCAAAGCAGGGGGCCGCCACTGCGGGCGACGGCGGGCAGCCCCACA

P D P S A S Y A F E L L G G F V R P A V T A A A A V L G V

TCCGCGGACCGGGCGCGGACCTTCGCGGATCTGCTGGAGCGGCTCCGGCCGCTGTCCGACAGCCTGCTGGCCCCCGCAGTCCCTGCGGAC
AGGCGCCTGGCCGCGCGCCTGAAGCGCCTAGACGACCTCGCCGAGCGCGGACAGGCTGTCCGACGACCGGGGCGCTCAGGGACGCCTG

P A D R R A D F A D L L E R L R P L S D S L L A P Q S L R T

FIG. 32D

119/164

sugar.finalgene b-1 Sequence

10	20	30	40	50	60	70	80	90
1234567890123456789012345678901234567890123456789012345678901234567890								
GGTACGGGGCGGACGGCGGCTGGCCGAGCTACGGCGCTGCTCGCGCTCCTCGGCAATCCGAGCAGTGGCGGGAGCTGTGCGACCGGCCCGG								1530
CCATGCCCGCCGCTGCCGCGGACCGGCTCAGTGCCTGCGGACGAGCGGCTAAGCTGTGAGGGGGCCCGGGACGACAGCCGCGAGCC								
V R A A D G A L A E L T A L L A D S D S P G A L L S A L G								
				TfiI			Apal	
				▼			▼	
BstEII								
▼								
GGTCACCGCAGCCGTCCAGCTCACCGGAGACCGCGGTGCTCGCGCTCCTCGGCAATCCGAGCAGTGGCGGGAGCTGTGCGACCGGCCCGG								1620
CCAGTGGCGTCCGACGGTCGAGTGGCCCTTTCGCCACGAGCGCGAGGAGCGCTAGGGCTCGTACCGCCCTCGACACGCTGGCCCGGCC								
V T A A V Q L T G N A V L A L A L A H P E Q W R E L C D R P G								
NotI								
▼								
GCTCGGGCGGCGGCGGAGAGACCTCCGCTACGACCCCGCGGTGCTGAGCTCGACCGCCCGGTGGTCCGCGGGAGACGGAGCTGGC								1710
CGAGCGCGCGCGCCACCTCCTCTGGGAGGCGATGCTGGCGGCGCACGTCGAGCTGCGGGCCACCCAGGCGCCCTCTGCCTCGACCG								
L A A A A V E E T L R Y D P P V Q L D A R V V R G E T E L A								
				NspHI			BbsI	Eco47III
				▼			▼	▼
GGGCGGGCGGCTGCCGGCGGGCGGATGTCGTCTGACCGCGGACCGCGCGGACCCGAGGTCTTACGGACCCCGGAGCGCTT								1800
CCCGCGCGCGACCGCGCGCGCGGTACAGCAGGACTGGCGGCGCTGGCGCGCCCTGGCCCTCCAGAAGTGCCTGGGCCCTCGCGAA								
G R R L P A G A H V V V L T A A T G R D P E V F T D P E R F								

FIG. 32E

120/164

sugar.final gene b-1 Sequence

10	20	30	40	50	60	70	80	90
1234567890123456789012345678901234567890123456789012345678901234567890								
RsrII ▼								
CGACCTCGCGCGCCCGACCGCGCGCGACCTCGCGCTGCACCCCGCGGTCCGTACGGCCCGGTGGGTCCCTGGTCCGGCTTCAGGC								1890
GCTGGAGCGCGCGGGGCTGCGSGCGCGCGTGGAGCGCGACGTGGGGCGGCCAGGCATGCCGGGCCACCGCAGGGACCGAGCCGGAAGTCCG								
D L A R P D A A H L A L H P A G P Y G P V A S L V R L Q A								
RsrII								
GGAGGTGCGCGTGGGACCCCTGGCCGGGGGTTCCTCCGGGTGGGGCAGCGGGGGACGTGCTCCGCCCCCGCGCGCGCTGTGCGCGG								1980
CCTCCAGCGCGACGCTGGACCGCGCCCGCAAGGGGCCCGACGCGCTCCGCCCCCTGCACGAGCGGGGGCGCGCGGACAGCCCGGC								
E V A L R T L A G R F P G L R Q A G D V L R P R A P V G R								
EspI Bpu1102I RsrII								
CGGGCCGCTGAGCGTCCCGGTGAGCGAGCTCCTGAGACACCGGGGGCCCGGTCCGTCGCGCCCGCGCCCTTCGGACGGACCGGACGGCTCGGAC								2070
GCCCCGCACTCGCAGGGCCAGTCGTCAGGACTCTGTGCCCCCGGGGCCAGGGCGGGGGGGAAGCCTGCCTGGCCCTGCCGAGCCTG								
G P L S V P V S S S								
RsrII								
CACGGGACGGCTCAGACCGTCCCGTGTGTCCCGGTCCGGCTCCCGTCCGCCCCCATCCCGGCCCTCCACCGGCAAGGAAGGACACGACGC								2160
GTGCCCCCTGCCGAGTCTGGCAGGGCACACAGGGGCAGGCCGAGGGCAGGGGTAGGGCGGGGAGGTGGCCGTTCCTTGCTGTGCTGCG								

FIG. 32F

121/164

[illegible]

FIG. 32G

122/164

sugar.finalgene b-1 Sequence

10 20 30 40 50 60 70 80 90
123456789012345678901234567890123456789012345678901234567890

CCGCTCCTGGCAGCCGGACCTGGTGTGTGGAGCCGACGACCTACGCGGGGCGCGTCCGCCGCCAGGTCAACGGTGCCGCCGACGCCG
GGCAGGACCGTCGGCCTGGACACGACACCCCTCGGCTGCTGGATGCGCCCGCGGACGCGCGGGTCCAGTGGCCACGCGCGCTGCGGGC
R S W Q P D L V L W E P T T Y A G A V A A Q V T G A A H A R 2610

AgeI
BstEII

Apal

GGTCCTGTGGGGCCCGACGCTGATGGCAGCGCCCGCCGCAAGTTCTGTCGGTGGGACCGGACCGCGCGCGGTCCAGTGGCGGCTCCTGGG
CCAGGACACCCCGGGCTGCACTACCCGTCGCGGGCGCGGTTCAAGCAGCGCGACGCCCTGGCCGTCGCGGGGCTCGTGGCGCTCCTGGG
V L W G P D V M G S A R R K F V A L R D R Q P P E H R E D P 2700

AgeI BstBI
EarI PvuI

CACCGCGAGTGGCTGACGTGGACGCTCGACCCGGTACGGGCGCTCCTTCGAAAGAGAGCTGCTCACCGGCCAGTTCACGATCGACCCGAC
GTGGCGCCTCACCGACTGCACCTGCGAGCTGGCCATGCCGCGGAGGAAGCTTCTCCTCGACGAGTGGCCGGTCAAGTGTAGCTGGGCTG
T A E W L T W T L D R Y G A S F E E L L T G Q F T I D P T 2790

CCGCGGAGCCTGGCCCTCGACACGGGCGCTGCCGACCGTCGGGATGCGTTATGTTCCGTACAACGGCACGTCGGTGTGCCGACTGGCT
GGGCGGCTCGGACGCGGAGCTGTGCCGAGCGCTGGCAGCCCTACGCAATACAAGGCATGTTGCCGTGACGCCAGCACGGCCTGACCGA
P P S L R L D T G L P T V G M R Y V P Y N G T S V V P D W L 2880

FIG. 32H

123/164

sugar.finalgene b-1 Sequence

10	20	30	40	50	60	70	80	90
123456789012345678901234567890123456789012345678901234567890								
GAGTGAGCCCGCCCGCGCCCGGGTCTGCCCTGACCCCTCGGCGTCTCGGCGGCGACGGCGTCTCGCAGGGCGA								2970
CTCACTCGGGCGGGCGCGCCCGGGCCAGACGGACTGGGAGCCCGCAGAGCGCGCACTCCAGGAGCCCGCGCTGCCGCGAGAGCGTCCCCGCT								
S E P P A R P A R P R V C L T L G V S A R E V L G G D G V S Q G D								
CATCCTGGAGGCGCTCGCCGACCTCGACATCGAGCTCGTGCCACGCTCGACGCGAGTCAGCGCGCCGAGATCCGGCAACTACCCGAAAGCA								3060
GTAGGACCTCCGCGAGCGGCTGGAGCTGTAGCTCGAGCAGCGGTGCGAGCTGCGCTCAGTCGCGCGGCTCTAGGCGTTGATGGGCTTCGT								
I L E A L A D L D I E L V A T L D A S Q R A E I R N Y P K H								
CACCCGGTTCACGGACTTCGTGCCGATGCA CGGCTCCTGCCGAGCTGCTCGGCGATCATCCACCGCGGGCGGGCACCTAGCGGAC								3150
GTGGGCCAAGTGCCCTGAAGCACGGCTACGTGCGCGAGGACGGCTCGACGAGCCGCTAGTAGTGTCGCCGCCCGCGCGTGGATGCGCTG								
T R F T D F V P M H A L L P S C S A I I H H G G A G T Y A T								
BclI								
CGCCGTGATCAACGGGTGCCGAGGTGATGCTGCCGAGCTGTGGACGCGCCGGTCAAGGCGGGCGCGTCCGAGCAGGGGGCGGG								3240
GCGGCACTAGTTGCGCCACGGCGTCCAGTACGAGCGGCTCGACACCTGCGGCGCCAGTTCCGCGCCCGCGCAGCGGCTCGTCCCCCGCCC								
A V I N A V P Q V M L A E L W D A P V K A R A V A E Q G A G								

FIG. 32I

124/164

sugar.finalgene b-1 Sequence

10	20	30	40	50	60	70	80	90
1234567890123456789012345678901234567890123456789012345678901234567890								
GTTCCTCCTGCGCGCGCGGAGCTCACGCCCGCAGGCGGTGCGGGACGCCCGTCTCCGCATCCTCGACGACCCCTCGGTGCGCCACCGCCGC								
CAAGAAGGACGCGCGCGCGGCTCCGAGTGCAGTCCGGCACGCCCTGCGGCAGCAGCGGTAGGAGCTGCTGGGGAGCCAGCGGTGGCGGCG								
F F L P P A E L T P Q A V R D A V V R I L D D P S V A T A A								
3330								
GCACCGGTGCGCGAGGAGACCTTCGGCGACCCACCCCGCGGGATCGTCCCGAGCTGGAGCGGCTCGCCGCGCAGCACCCGCCGCC								
CGTGGCCGACGCGCTCCTTGGAAGCCGCTGGGGTGGGCGCGCCCTAGCAGGGGCTCGACCTCGCCGAGCGCGCGTCTGTTGGCGCGGG								
H R L R E E T F G D P T P A G I V P E L E R L A A Q H R R P								
3420								
StuI								
GCCGGCCGACGCCCGCGCTGAGCCCGCACCCCTCGCCCGCAGGCTCACCCCTGTATCTGCGCCGGGGAGCGCCCGGCCACCCCTCCGA								
CGGCCGCTGCGGGCGGTGACTCGGCGTGGGAGCGGGGTCCGGAGTGGGACATAGACGCGGCCCCCTGCGGGGCGCGGTGGAGGCT								
P A D A R H								
3510								
AatII								
EarI								
AAGACCGAAAGCAGGAGCACCGTGTACGAAGTCGACACCGCGACGTCTACGACCTCTTCTACCTGGGTGCGGGCAAGGACTACGCCGCC								
TTCTGGCTTTCGTCTCGTGGCACATGCTTCACTGGTGGGCTGCAGATGCTGGAGAAGATGGACCCAGCGCCGTTCTCTGATGCGGCGG								
V Y E V D H A D V Y D L F Y L G R G K D Y A A								
3600								

FIG. 32J

125/164

sugar.finalgene b-1 Sequence

10 20 30 40 50 60 70 80 90
123456789012345678901234567890123456789012345678901234567890

StuI

StuI

GAGGCTCCGACATCGCCGACCTGGTGGCTCCCGTACCCCGAGGCCTCCTCGCTCCTGGACGTGGCCTGCGGTACGGGCACGCATCTG
CTCCGGAGGCTGTAGCGGCTGGACCAACGCGAGGGCATGGGGCTCCGGAGGAGCGAGGACCTGCACCGGACGCCATGCCGTGCGTAGAC
E A S D I A D L V R S R T P E A S S L L D V A C G T G T H L 3690

StyI

NspHI

GAGCACTTCAACCAAGAGTTCCGGCAGACACCGCCGGCTGGAGCTGTCCGAGGACATGCTCACCCACGCCCGCAAGCGGCTGCCCGACGCC
CTCGTGAAGTGGTTCTCAAGCCGCTGTGGCGGCCGACCTCGACAGGCTCTGTACAGTGGGTGGCGGCTTCGCCGACGGGCTGCGG
E H F T K E F G D T A G L E L S E D M L T H A R K R L P D A 3780

NspHI

NspHI

ACGCTCCACCGGACATCGCGGACTTCCGGCTCGGCCCGGAAGTTCTCCGCCGTGGTCCAGCATGTTCAAGCTCCGTCCGTACCTGAAG
TGCAGGTGGTCCCGCTGTACGCCCTGAAGCCGAGCCGCCCTTCAAGAGCGGCCACCAAGTCCAGTCCAGGCGAGCCGATGGACTTC
T L H Q G D M R D F R L G R K F S A V V S M F S S V G Y L K 3870

BbsI

ACGACCGAGGAACTCGGCGCGCGCTCGCCTCGTTCCGCGGAGCACCTGGAGCCCGGTGGCGTCTCGTCCGAGCCGTGGTGTCCCG
TGCTGGCTCCTTGAGCCGCGCGGAGCAAGCGCCTCGTGGACCTCGGGCCACCGCAGCAGCAGCTCGGCACCAAGGCGC
T T E E L G A A V A S F A E H L E P G G V V V E P W F P 3960

FIG. 32K

126/164

sugar.finalgene b-1 Sequence

10 20 30 40 50 60 70 80 90
123456789012345678901234567890123456789012345678901234567890

AatII ▼

DraIII ▼

GAGACCTTCGCCGACGGCTGGGTCAAGCCCGACGTCCTCCCGCTGACGGGCGCACCGTGCCCGCTGTCTCGCACCTCGGTGCGGAGGGG
CTCTGGAAGCGGCTGCCGACCCAGTCGCGGCTGCAGCAGGCGGCACTGCCCGCTGGCACCGGGCACAGAGCGTGAGCCACGCCCTCCCC
E T F A D G W V S A D V V R R D G R T V A R V S H S V R E G 4050

AatII ▼

AACGCCAGCGGCATGGAGTCCACTTACCGTGGCCGACCCGGGCAAGGGCGTGCGGCACCTTCCGACGTCCTCATCACCCCTGTTC
TTGCGCTGCGGCTACCTCCAGGTGAAGTGGCACCGGCTGGGCCCGCTTCCCGCACGCCGTGAAGGCTGCAGGTAGTAGTGGGACAAG
N A T R M E V H F T T V A D P G K G V R H F S D V H L I T L F 4140

SfiI ▼

EcoRI ▼

CACCAGGCCGAGTACGAGGCCGCGTTACGGCCCGCGGCTGCGGTCGAGTACCTGGAGGGCGGCCCGTGCGGCGCTCTTCGTC
GTGGTCCGGCTCATGTCCGGCGCAAGTGCAGCGGCGCCGACGCGCAGCTCATGGACCTCCCGCGGCGAGCCCGGACCGGAGAGCAG
H Q A E Y E A A F T A A G L R A V E Y L E G G P S G R G L F 4230

AatII ApaI ▼

GGCGTCCCGGCTGAGCACCGCCCAAGACCCCGCGGCGGACGTCCCGGCTGCACCAAGCAAGAGAGAGGAGAACCGGTGACAGGT
CCGAGGGGCGGACTCGTGGCGGTTCTTGGGGGGCCCCCGCCCTGCAGGGCCACGTTGTTCTCTCTCTTGTGCTTGGCACTGTCCA
G V P A V T G 4320

FIG. 32L

127/164

sugar.finalgene b-1 Sequence

10 20 30 40 50 60 70 80 90
1234567890123456789012345678901234567890123456789012345678901234567890

NotI
▼

AAGACCCGAATACCGCGTGTCCGCGCGCGCCGACACGCCCCAGGGCCCTTCAACCCTGGCCGTCGTGGCACCCCTGCTGGCGGGCACCACC
TTCCTGGGCTTATGGCGCACAGGCGGCGCGCGCTGGTGGGAGTCCGGAACTGGGACCGGCGCAGCAGCCGTGGGACGACCGCCCGTGGTGG
K T R I P R V R R G R T T P R A F T L A V V G T L L A G T T 4410

NotI
▼

GTGGCGGCGCGCTCCGCGCGCGCCGACACGGCCCAATGTTCACTACACGAGCCGGCGGGAGCTCGTCGCCACAGATGACGCTCGAC
CACCGCGCGCGAGGCGCGCGCTGTGCCGGTTACAAGTCACTGTGCTGGCCCGCCCTCGAGCAGCGGGTCTACTGCGAGCTG
V A A A P G A A D T A N V Q Y T S R A A E L V A Q M T L D 4500

GAGAAGATCAGCTTCGTCCACTGGGCGCTGGACCCCGACCGGCGAGAACGTCGGCTACCTTCCCGCGGTGCCGCTGTGGCATCCCGGAG
CTCTTCTAGTCGAAGCAGGTGACCCCGGACCTGGGGCTGGCCGCTCTTGACGCCGATGGAAAGGCGCGCACGCCGTAGGGCCTC
E K I S F V H W A L D P D R Q N V G Y L P G V P R L G I P E 4590

MscI
BamI
▼

CTGCGTCCGCGCGACGGCCGAAACGGCATCCGCTGTGGGGCAGACCGCCACCGGCTGCCCGCGCGGTCCGCTGGCCAGCACCTTC
GACGCACGGCGCTGCCGGCTTGCCTAGGCGGACCAACCCGCTGTGGCGGTGGCGGACCGGCGCGGCGGACCGGTCTGTGAAG
L R A A D G P N G I R L V G Q T A T A L P A P V A L A S T F 4680

FIG. 32M

128/164

[illegible]

pflMI ▼ **Apal** ▼

GACGACACCATGGCCGACAGCTACGGCAAGTCAATGGGCCGCGACGGTCGGCGCTCAACACGAGCATGGTCTCTGGGCCCGATGATGAAC
CTGCTGTGGTACCGGCTGTGATGCCGTTCCAGTACCCGGCGCTGCCAGCGCGGAGTTGGTCTGTACGAGGACCCGGGCTACTACTTG
D D T M A D S Y G K V M G R D G R A L N Q D M V L G P M N

AAACATCCGGGTCCGCACGGCGGCCGGAACATCGAGACCTTCAGCGAGGACCCCTGGTCTCCTCGGCAACCGCGGTGCCCCAGATCAAG
TTGTAGGCCCAACGGCGTGCAGCCGGCCCTTGATGCTCTGGAAGTCTCGTGGGGACCAAGAGCGCGGTGGCGCCAGCGGGTCTAGTTC
N I R V P H G G R N Y E T F S E D P L V S S R T A V A Q I K

GGCATCCAGGTGCGGTCTGATGACCACGGCCAAAGCACTTCGGGCCAAACAACGAGAGAAACAACCGCTTCTCCGTGAACGCCAATGTC
CCGTAGGTCCACGCCCAGACTACTGGTGCCGGTTCGTGAACGCCCGGTGTGTGGTCTCTTGTGGCGAAGAGGCACCTTGGCGTTACAG
G G I Q G A G L M T T A K H F A A N N Q E N N R F S V N A N V

styI sfil

GACGAGCAGACGCTCCGGAGATCGAGTTCCCGCGCTTCGAGGGCTCTCCAAGCCGGCGGCTCCTTCATGTGTGCTTACACGGC
CTGCTCGTCTCGAGGGCTTAGCTCAAGGGCGCAAGCTCCGACGAGGTTCCGGCCCGCGCGGAGGAAGTACACAGGATGTTGCCG
D E O T L R E I E F P A F E A S S K A G A A S F M C A Y N G

FIG. 32N

129/164

sugar.finalgene b-1 Sequence

10 20 30 40 50 60 70 80 90
123456789012345678901234567890123456789012345678901234567890

FspI DraIII
▼ ▼

CTCAACGGGAAGCCGTCCTGCGGCAACGACGAGCTCCTCAACACGTGCTGCGACGACGTGGGCTTCCAGGGCTGGGTGATGTCCGAC
GAGTGGCCCTTCGGCAGGACGCCGTTGCTGCTCGAGGAGTTGTTGCACGACGCGTGCCTACCCCGAAGGTCCCGACCCACTACAGGCTG
L N G K P S C G N D E L L N N V L R T Q W G F Q G W V M S D 5130

StyI EcoNI PflMI AatII
▼ ▼ ▼ ▼

TGGCTCGCCACCCCGGCAACGACGATCGACCAAGGCGCTCGACAGGAGATGGGCGTGGAGTCCCGGCGACGTCCCGAAGGCGGAG
ACCGAGCGGTGGGCGCCGTTGGTTCCTCGGAGCTGGTCTCTACCCGACGCTCGAGGGGCGGCTGCAGGGCTTCCCGCTC
W L A T P G T D A I T K G L D Q E M G V E L P G D V P K G E 5220

BbsI
▼

CCCTCGCCCGCGCCAAAGTTCTTCGGCGAGGCGCTGAAGACGGCCGTCCTGAACGGCACGGTCCCGAGGGCGCGTGACGCGGTCCGGCG
GGGAGCGGCGCGCGTTCAAGAAGCCGCTCCGCGACTTCTGCGGCGAGGACTTGCCGTGCCAGGGGCTCCCGCGGCACTGCGCCAGCCGC
P S P P A K F F G E A L K T A V L N G T V P E A A V T R S A 5310

GAGCGGATCGTCGGCCAGATGGAGAAGTTTCGGTCTGCTCCTCGCCACTCCGGCGCGCGGCGCGGCGGACGAGCGGGTGCCACAGGCG
CTCGCCTAGCAGCCGGTCTACCTCTTCAAGCCAGACGAGGAGCGGTGAGGCCGCGGCGCGGCGTGGTTCGCGCTGTTCGCGCCACGGGTCCGC
E R I V G Q M E K F G L L L A T P A P R P E R D K A G A Q A 5400

FIG. 320

130/164

[illegible]

FspI **AlwNI**

GTGTCCCGCAAGGTCGCCGAGAACGGCGCGGTCTCTTGGCGCAACGAGGGCCAGGCCCTGCCGCTCGCCGGTGACGCCGGAAGACATC
CACAGGGCGTTCACGGGCTCTTGGCGCGCCACGAGGACGGCTTCTCCGTCGGGACGGCGAGCGGCCACTGCGGCCGTTCTCTCGTAG
V S R K V A E N G A V L L R N E G Q A L P L A G D A G K S I

Styl **BstEII**

5580
GCGGTCA TC GGGCCGACG GCGCTCGA CCGGCTCGG CCGCCACG TCGTCCCGACTCG GCGGGCGGCCACTCGAC
CGCCAGTAG CCGGGCTG CCGGCAGCTGGGTTCCAGTGGCCG GACCCGTCGCGGGTGCAGCAGGCCCTGAGCCGCCCGGGTGAGCTG
A V I G P T A V D P K V T G L G S A H V V P D S A A P L D

5670
ACCATCAAGGCCCGCGGGTGCGGTGACGTACGAGACGGGTGAGGAGACCTTCGGACGCGAGATCCCGCGGGGAACCTC
TGGTAGTTCCGGGCGCGCCACGCCACGCTGCCACTGCTCTGCCACTCTCTGGAAGCCCTCGGTCTAGGGCCGCCCTTGGAG
T I K A R A G A G A T V T Y E T G E E T F G T Q I P A G N L

XhoI
PaeR7I

AGCCCGGGCTTCAACAGGGCCACCAGCTCGAGCCGGGCAAGGGGGGGCGCTGTACGACGGCACGCTGACCGTGCCCCCGCCGACGGCGAG
TCGGGGCCGCAAGTTGGTCCGGTGGTCGAGCTCGGCCCGCTTCGGCCCCCGGACATGCTGCCGTGCGACTGGCACGGGGCGGTGCCGCTC
S P A F N Q G H Q L E P G K A G A L Y D G T L T V P A D G E

FIG. 32P

[illegible]

FIG. 32Q

sugar.finalgene b-1 Sequence

[illegible]

FIG. 32R

133/164

sugar.finalgene b-1 Sequence

10	20	30	40	50	60	70	80	90
123456789012345678901234567890123456789012345678901234567890								
ACGTCCACGGGTGCTGAAGTCAAGGTACGGTCAAGTCCGCAACAGCGGGAAGCGCGCGCGGCGGAGGTCGTCCAGGCGTACCTCGGTGCC								6570
TGCAGGTGCCCCACAGACTTCCAGTGCCAGTGCAGGCGTTGTGCGCCCTTCGCGCGCGCGCGGTCTCCAGCAGGTCCGTCATGGAGCCACGG								
T S T G G L K V T V T V R N S G K R A G Q E V V Q A Y L G A								
AGCCCGAACGTGACGGCTCCGACGGCGAAGAAGTCTGTGGGTACACGAAGTCTCGCTCGCGCGCGGCGGAGGCGGAGCGGTGACG								6660
TCGGGCTTGCACTGCCGAGGCGTCCGCTTCTTCGAGCACCCGATGTGCTTCAGAGCGAGCGCGCGCGCTCCGCTTCTGCCACTGC								
S P N V T A P Q A K K K L V G Y T K V S L A G E A K T V T								
GTGAACGTGACCGCGGTGACGTGACACCGGTTCGTCTCCGCGGACCTGCGGGGAGCGCCACGGTCAACGTCTGGTGACGTGACGCC								6750
CAC TTGCAGCTGGCGGCAGTCGACGTCTGGCCAAAGCAGGAGGCGGCTGGACGCGCCCGGTCCAGTTGCAGACCACTGCACTGCGG								
V N V D R R Q L Q T G S S S A D L R G S A T V N V W								
GTGAAAGCGGCGGTGCCCGCCACCCCGGAGGTTGGCGGACCGCTTTTTCGGCCTGCTGGGTCTACCGGACCACTGACTAGGCCTGGT								6840
CAC TTTCGCGCCACGGGCGGTGGGCCCTCCACCGCCCGTGGCGAAAGCCGAGCACCCAGATGGCTGGTGGACTGATCCGGACCA								
A Q								

FIG. 32S

134/164

sugar.finalgene b-1 Sequence

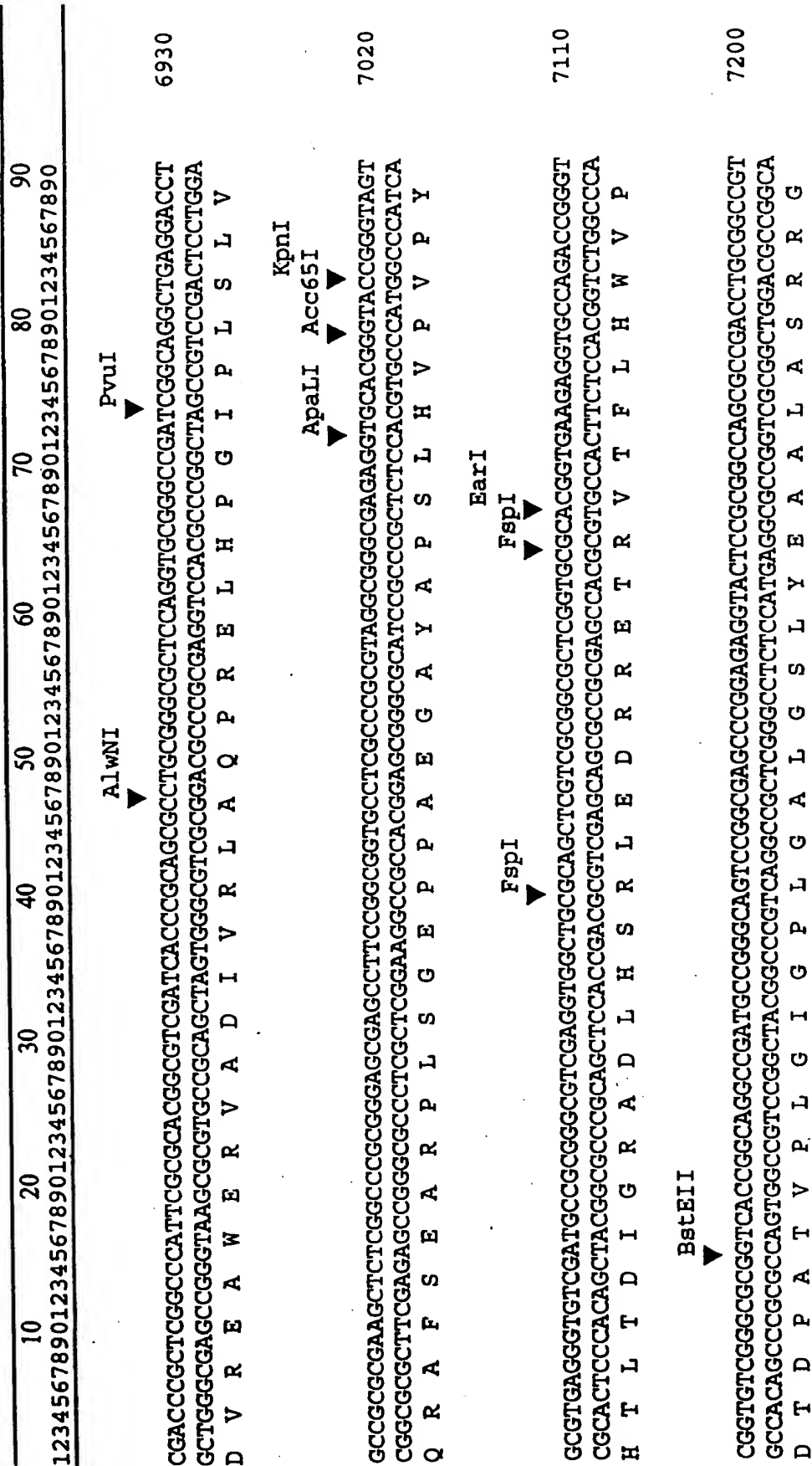


FIG. 32T

135/164

sugar.finalgene b-1 Sequence

10	20	30	40	50	60	70	80	90
123456789012345678901234567890123456789012345678901234567890								
PvuII		BamHI				XcmI		
▼	▼					▼		
TCCAGCTGTCCAGGTGGCGGAGCCGGATCCGACGACGGCGGCTTGCATCTCGTCCAGGGGGAGTTGGTGCCCTTCGTCTCGTGGCTGT								7290
AGGTCGACAGGTCCACCGCTCGGCTAGGCGTCTGTCGCCCGGACGTAGAGCAGGTCCGCCCTCAACACGGGAAGCAGAGCACCCGACA								
N W S D L H A L R I R L V A A Q M E D L R S N T G K T E H S								
			BspMII					
			BspEI					
			▼					7380
ACTTCTGCCGCGAGCCGTAGTTGCGGAGCATCCGGAGCCGTTCCGGGAGCTCGGGTCCGCCGTGACGACGGCGCCCGCTCGGCCGAAGC								
TGAAGACGGCGCTCGGCATCAACGCCTCGTAGGCCTCGCAAGCCGCTCGAGCCCGGCTGCTGCCGCGCGGCGGACGGGCTTCG								
Y K Q R S G Y N R L M R L R E A L E P D G T V V A G G D G F								
								7470
AGCCGAGGTTCTTGCCCGGTAGAGCTGAACGGCGGCCACCGACGACCCGGCGCCGATCCGCCCGCCCGGTAGCGGGCGCCGTGGGCCT								
TCGGCTCCAAGAACGGGCCCATCTTCGACTTGGCCCGGTGGCTGCTGGGCCCGGCTAGGCGGCCGCCATCGCCCGCGGCACCCCGGA								
C G L N K G P Y F S F A A V S S G A G I R R G R Y R A G H A								
								7560
			RsrII					
			▼					
GCGGGCGTCTCGACGATGTGCAGGCGCGTGCCTGCGGAGCTCGCGGAGGGCGTCCATGTCCGGGGTGCCCGTAGAGTGGACGG								
CGGCGCGCAGGAGCTGTACAGTCCGGCAGCGCCAGGCGCTCGAGCGCCTCCCGCAGGTACAGCCGCCCCACGGGCATCTCCACCTGCC								
Q A A D E V I H L G H R D A L E R L A D M D A P H G Y L H V								

FIG. 32U

136/164

sugar.finalgene b-1 Sequence

10 20 30 40 50 60 70 80 90
123456789012345678901234567890123456789012345678901234567890

GGAGGAGCGCCCGGGTGGGGGGTGATCGCCTTCTCGACGAGCAGCGGGTCCAGGGTGGGGTGGTCTCTGTGCGGCTCGACGGGCAAGG
CCTCCTCGCGGGCCACGCCCCCCTAGCGGAAGAGCTGCTCGCCAGGTCCACCCACAGGAGCACGCCGAGCTGCCGTGCC
P L L A R T R P T I A K E V L L P D L T P H D E H P E V P V 7650

BsaAI
AflIII
PvuII

GGGTCCGCGCGGTGGCGGACACCGGAGCCAGCTGGCGATGTACGTGTCGAGGGGACGATCACCTCGTCCCCTGGGTCCGATGCCGAGGC
CCCAGCGGGCCACCGCCTGTGGCGCTCGGTGACCGCTACATGCACAGCTCCCTGTCTAGTGAGCAGGGGCCAGGCTACGGCTCCG
P T A G T A S V A L W S A I Y T H S P V I V E D G P G I G L 7740

TfII

CGCGGAGGGCGAGCTGGAGGGGTCCATCCCGTGTTCACGCCGACGGCGTGGTCCGTCTCGCAGTACGCGGCGAACTCCGCCCTCGAATC
GCGCCTCCCGCTCGACCTCCCGCAGGTAGGGCGACAAGTGGCGTGCAGCACGAGCGTATGCGCCGCTTGAGCGGAGCTTAG
G R L A L Q L A D M G S N V G V A H D T E C Y A A F E A E F 7830

PvuI

CTTCGAGTTCGGGTCCGAGGAGGTAGCGCCCGAGTTCGAGGACGCGGGCGATCGCGCGTCTCGCGCGGAGCTCCTCGTAGGCGG
GAAGCTCAAGCCCGAGGCTCCTCCATCGCGGGGCTCAGCTCCTGCGCCCGTAGCGCCGAGCCAGAGCGCGCTCGAGGAGCATCCGCC
G E L E P G L L Y R G S D L V R A I A A D T E A R L E E Y A 7920

FIG. 32V

137/164

[illegible]

KpnI
 Acc65I RsrII
 ▼ ▼
 CGCCTTGAGCGGCTCCACACAGCCGCGTTCCTCCGCGAGGCGTCCGCGAAGAGACCTCGGGCGGTA
 GCGGAACCTCGCCGAGGGTGGTCGGCGCCAAAGAGGGCCATGGTCGCCCTGCCAGGCGCGCTCCGGCAGGCGCTTCTCTGGACGCCCGCCAT
 A K L P E W G R N E R Y W R V T R A L G D A F S V Q P R Y
 RsrII

BspMII
 BspEI
 ▼ ▼
 GCGGAGCTCGCGCTCGATCTCGCCGCGCTCGAGGAGTAGCGCAGGTCTGTGGCCCTTGCGGTCCGGACCGAGGACCAGTC
 CGGCTCGAGCGCGAGCTAGAGCGGGCGGAGCTCCCTCATCGCGTCCAGCACCGGGAACGCCAGCCGCTGGAAGCCCTGGCTCCTGGTCAG

BglII
▼
GGCGCCGAGCGAGTCCAGGAGGATGCCGGTGGTCAGCTCCAGGCCGCCGATGTGGTAGATCTCGCCGGCCCGGCC
CCGGGCTCGCTCAGGTCTCTACGGCCACTCAAGGCCAACCAAGTCGAGGTCCGGGGCGGCTACACCATCTAGAGCGGCCGGGCCGG
A G L S D L L I G T L E R N T L E L G G I H Y I E G A R G

FIG. 32W

138/164

sugar.finalgene b-1 Sequence

10	20	30	40	50	60	70	80	90
1234567890123456789012345678901234567890123456789012345678901234567890								
G C C C G C G A G A C G A G C G C C C C G C A G T G G T C G T C G G T G T G C A C C C A C T C G C G G A C G T T C G C C C G T C G C C G T A C A G C G G A G C G T								
C G G C G C T C C T G C T C G C G T A C G G G C C G T C A C A G C A G C C A C A C G T G G T G A G C C C T G C A A G C G C G G C A T G T C G C C C T C G C A								
G A L V L A I G R C H D D T H V W E R V N A G D G Y L P L T								
A p a L i								
G C C G C C G T C G A G A G G T T C G T C A C G A A G A G G G G A T G A G C T T C T C G G G G T G T G T A C G G C C C G T A G T T G T T G C A G C A G C G G T G A T C C G								
G G C G C G C A G C T C C T C C A A G C A G T G C T T C C C C C T A C T C G A A G A G C C C C A C G A C C A T G C C G G G C A T C A A C A C G T C G T C G C C C A C T A G G C								
G G D L L N T V F L P I L K E P H Q Y P G Y N N C C R T I R								
E a r I								
T A C G T C G A G G C C G T A C G T C C G G T G G T A G C C G G G C A A C G A G T C G G A G C C G G C C T T G G A C G C C C G T A G G G C G A G T T G G G C T C C A G C G G								
A T G C A G C T C C G G C A T G C A G G C C A C C A T C C G G C C C G T T G C T C C A G C C T C G G C C G G A A C C T G C G G C G C A T C C C G C T C A A C C C G A G G T C G C C								
V D L G Y T R R H Y A R A V L D S G A K S A A Y P S N P E L P								
S t y I								
G C T G C T C G G T C C A G A G C C G G A G T C G A T C G A C C C G T A C A C C T C G T C G G T G G A G A C G T G C A C G A C C C G G C C G A C G C C G G C G T C G A C G G C								
C G A C G A G A G C C A G G T C C T C G G C C T C A G C T A G T G G G C A T G T G A G A G C C A C C T C T G C A C G T G C T G G G C C G G C T G C G G C C G C A G C T G C C G								
S S E T W S G S D I S G Y V E D T S V H V V R G V G A D V A								
P v u I								
A p a L i								

FIG. 32X

139/164

sugar.finalgene b-1 Sequence

10	20	30	40	50	60	70	80	90
1234567890123456789012345678901234567890123456789012345678901234567890								
AlwNI ▼								
GCACTGGAGCAGCGTCTGCGTGGCCCTGCAAGTTGGTCTCGGTGAACACGACGCGCCCGCGATGGAGCGGTCCACGTTGGCTCTCGGCCCGC								
CGTGACCTCGTCGACAGCAGGGACGTGCAACACGAGCCACTTGTGCTGCGCGGCGCTACCTCGCCAGGTGCACCGAGAGCCCGCG								
C Q L L T Q T G Q V N T E T F V S A G A I S R D V H S E A A								
PmlI BsaAI ▼								
8730								
GAAGTGGACGATGGCGTCCACGCCCGCGCAGTTCCCGGGCGAGAGCGCGGCGTGGCGATGTGCGCGTGGACGAAGCGCAGTCGCGGGTC								
CTTACCTGCTACCGCAGGTGCGCGCGTCAAGGGCCCGCTCTCCGGCCGACGCGCTACAGCGCACCTGCTTCGGTTCAGCGCCCGCAG								
F H V I A D V G R L E R A L L G A D R I D G H V F R L R P D								
8820								
CGGTCCACCGGGCGAGGTTGGCGCGGTTGCCCGCGTAGGTGAGGCTGTCCAGGACGATCACCTCATCGCGGGCACGTCGCGGTACGC								
GCGCAGGTGGCCCCCGCTCCAACCGCGCCCAACGGGGCGCATCCACTCCGACAGGTCCTGCTAGTGGAGTAGCCCGCCCGTGCAGCCCCCATGCG								
A D V P A L N A R N G A Y T L S D L V I V E D A P V D P Y A								
8910								
DraIII ▼								
BstEII BbsI ▼								
9000								
CCGGCGAGGAGCTGCCGACGAAGTGGAGCCGATGAAGCCCGCACCTCCGCTACCCAGAACCGCACCTGCCGTCTTCCTTCGGTCCG								
GGCGCGTCTCTCGACGGCGTGTTCACGCTCGGCTACTTCGGGCGTGGAGGCCAGTGGTCTTCGGCGTACGGCAGGAGGAAAGCCAGCG								
G A L L Q R V F H S G I F G A G G T V L L R V								

FIG. 32Y

[illegible]

FIG. 32Z

141/164

sugar.finaIgene b-1 Sequence

10	20	30	40	50	60	70	80	90
123456789012345678901234567890123456789012345678901234567890								
AatII								
▼								
GCCGGCCCGCTCAGGTAGACGGGTTGACGTCGGTGATCTCCAGCTCGCCGGCGGCGAGGGCCGGATGTTCTTGGCGATGTCGACG								9450
CGGGCCGGGGGAGGTCCATCTGCGCCAACTGCAGCCACTAGAGTCCAGCGGCGCGCTCCCGGCTACAAGAACCGCTACAGCTGC								
A R G R E L Y V R N V D T I E L E G R P S P R I N K A I D V								
AatII								
▼								
ACGTCGTTGTCGTAGAGGTAGAGCGCGGTGACGGCGAGGTTGGAGCGCGCTTGACGGGCTTCTCGACGAGGTCGGTCAGCCGGCCCGTC								9540
TGCAGCAACAGCATCTCCATCTCCGGCCACTGCCGCTCCAACCTCGCGCCGAAGTCCGAGAGTGTCTCAGCCAGTCCGGCCGGGCGAG								
V D N D Y L Y L G T V A L N S R P K V P K E V L D T L R G T								
EarI								
▼								
GCGTCCACCTCGGGCAGCGCCGTACCGCTCGGGTCTTGTGACCGGGTAGCCGAAGAGCACGACGCCGTCGAGGGCGCGGATGCTGTCCCGC								9630
CGCAGGTGGAGCCGCTGCGGCATGGCGAGCCCCAGGAACCTGGCCCATCGGCTTCTCGTGGCTCGGCAGCTCCGCGGCTACGACAGGGCG								
A D V E A V G Y R E P D K V P Y G F L V C G D L R A I S D R								
ApaI								
▼								
AGGAGCGTGTAGAGCGCGGGCCCGTGGAGATGTTGTGCGCCAGGATCAGGGCGCAGGTGTGTCGCCGATGTGCTCGGCTCCGACGAGA								9720
TCCTCGCACATCTCCGGCCCGGCGACCTTCTACAACAGCGGGTCTTAGTCCCGGCTCCACAGCGGCTACACGAGCCGAGGCTGCTCT								
L L T Y L G P G H F I N D G L I L A C T D D G I H E A G V L								

FIG. 32AA

142/164

sugar.finalgene b-1 Sequence

10	20	30	40	50	60	70	80	90
1234567890123456789012345678901234567890123456789012345678901234567890								
TfII								
▼								
AGTGGTCCGCGATTCTCTGGACCGCATAGTCGAGTTCTATCCAGGTCCCTGCCGTTTCCGAGAACGCACTGGAAG								9810
TCACGCGCGGCTAAGGACGCCCGAGAAAGACCTGGCGTATCAGCTCAAGATAAGGGTCCACGACGGCAAGGCTCTTCGCTGACCTTC								
L A D A I G A P E K Q V A Y D L E I G L H R G N G L L S Q F								
TfII								
▼								
AGTTCGATGTCTGGGGGTCGAGATGATTTGAATCTCGGAATACCGCCGAGCATGAGAACCGACGCGGATAGATCATCGGTTTG								9900
TCAAGCTACAGACCCCCAGCTCTACTAAACTTAGAGCGCTTATGGCGGCTCGTACTCTTGGCTGTGCGCTATCATCTAGTAGCCAAAC								
L E I H Q P T S I I Q I E R I G G L M L V S L P Y Y I M P K								
TfII								
▼								
TTGTAGACCGGAAGAAATCTGCTTCGAAATGACCGAGGTCCCGGATGCAGCCGAGTCCGCTCCCGCCGCGGCACTATTCCCTTCATT								9990
AACATCTGGCCTTCTTAGACGAAGCTTTACTGGCTCCAGCGGCTACGTCCGCTCAAGGCGAGGCGGCGGCTCCTGATAAGGGAAGTAA								
N Y V P L I Q K S I V S T A P H L R T G S G G A L V I G K M								
TfII								
▼								
CTCGGAAACTAGCAGCAGGGCGCGGTGATAACGGTCGGCGTGGCGAGTTAGGGGGGCGCTAGGGGCTGCGCAGGGGAGTGTACCCACC								10080
GAGCCTTTGATCGTCCCGCGGCCACTATTGCCAGCCGACCGCTCAATCCCCCGCGATCCCGACGCGTCCCCCTCACAGTGGTGG								

FIG. 32BB

[illegible]

FIG. 32CC

144/164

sugar.finalgene b-1 Sequence

10	20	30	40	50	60	70	80	90
1234567890123456789012345678901234567890123456789012345678901234567890								

10530
AGGAGATCGACCTGCACCGGACCTGCGGGTCCGCGGTGGGCGCATGGTGGGGGCTTGATCCGCAGCAGTTCCGGCGTCGGCCCCCGGTG
TCCTTAGCTGGACGTGGCGTGGACGCCCCACGCGCGTACCAACGCCCCGAACTAGGCGTCGTCAAGCCGAGCCCGGGGCCAC
L L D V Q V A V Q P H A T P R M T A P K I R L L E A D A G T

FspI



10620
CGCAGGCTGTTCAGGGCGTAGCCGCTAGTCGATGTGGAGTCCGGGGTGGCTCGCGGACCCGCTCCTCGAAGCGTTGAGGGCCTCCTGG
GGTCCGACAAGTCCCGCATCGGCATCAGCTACACCTCAGGCCCCCAGCGAGCGCCTGGCGGAGGAGCTTCCGCAACTCCCGGAGGACC
R L S N L A Y G Y D I H L G P T R E R V R E E F A N L A E Q

SfcI



NruI



PvuI



10710
AGTCGGCCCCGCTCCTCCTCGCGGAGCTTGCCGTCGTACGGCCGCTGTAGTCTCCTCGGAATGTTGACGAAGTCGATCGTCTGCCCTGC
TCGAGCCGGCGAGGAGGACGCCGTCGAACGGCAGCAGTGCCCGCGACATCAGGAGCGCTTACAACCTGCTTCAGCTAGCAGGACGGGACG
L E A R E E Q P L K G D D R G S Y D E R I N V F D I T R G Q

10800
CCGGCGTCGTTGAGTCGGCGATGAAGTCGACACGAGTCGAGCAGCGGGAGGACCGCCCCGGAGCAGCATGTAGCGGAAGCCGAGGTTG
GGCCCGCAGCAACTCCAGCCGCTACTTCAGCTGGTCCAGCTCGTCCGCCCTCCGTCCCGGCCCTCGTGGCTACATCCGCTTCGGCTCCAAC
G A D N L D A I F D V L D L L R S A R G P L V I Y A F G L N

FIG. 32DD

145/164

sugar.finalgene b-1 Sequence

10 20 30 40 50 60 70 80 90
123456789012345678901234567890123456789012345678901234567890

AlwNI
PvuII
▼ ▼

ATCGGCGACTCGCGCTCGGCGCGCAGCTGCTGGAAGCGGCGCAGGTTCTCGGGACGCGGCGGAAGGCGGCCTTCTTGCCTGGTCTGCTGC
TAGCCGCTGAGCGCGAGCGCGCGTCCGACGACCTTCGCGCGCTCCAGAGCGCTTCCGCGCGCTTCCGCGGGAAGAACGCGCCACGAGACG
I P S E R E A R L Q Q F R R R L N E R V R R F A A K K G T T Q 10890

SfiI
▼

TCGTACTCCTCGTTCGTTGAGCCCGTAGAGCGAGGTGCGGATGGCGTGCAGGCCCGGCGTGGCGCTCCAGGGTGCCTCGGTG
AGCATGAGGAGCAGCAACTCCGGCATCTCGCTCCACGCCTACCGCACGTCCGGGTCTCCGCGCGGCTCCGCGGAGTCCACGCGAGCCAC
E Y E E D N L G Y L S T R I A H L G W L G P Q R E L T R E T 10980

XmnI
▼

AGCGCGAAGGAGTTCGTGTAGACGGTGGGCGCGCAGCCCGTGGTGGGTGGCGGCGCAGGCTCCCGAGGCGCGGGTTGGTGAGCGGC
TCGCGCTTCCTCAAGCACATCTGCCACCCGCGTCCGGCACCCAGCCACCGCACCGCGCGGTCCGAGGGTCCGCGCCCAACCACTCGCCG
L A F S N T Y V T P R L G H D T A H A A L S G L G P N T L P 11070

TCCAGGCGCGGAGAGTACATCGCCGAGGGGTGCGCGGGGTATCTCGTGCATGACCGACCGGAACATGGCGTTGCCGGCGTCGAGG
AGGTCCGGGGCTCTTCATGTAGCGGCTCCCCAACGGGCGCCCATAGAGCAGTACTGGCTGGCTTGTACCGCAACGCGCCGAGCTCC
E L G G S F Y M A S P N G A P I E D I V S R F M A N G A D L 11160

FIG. 32EE

146/164

[illegible]

GCGGACGGGTCGTAGCGGGCGCGGTCAACGGACGCAGAACTGGCAGCGGAACATGCAGGTCGGCGCGGGGTAGAGGCCGACGCTGTAC
 CGCCTGCCACGATCGCCCGGCGCAGTGTCCTGCGTCTTCACCGTCGCTTGTACGTCCAGCCCGCCCCCATCTCCGGCTGGGACATG
 A S P D Y R A G T V R V C F H C R F M C T P G P Y L G V S Y

BibI

GGGAAGACGGGCTTCTTGGCGAGCGCCCGCTCGAAGACGCGCGCTGTTTCGAGCGGAGCAGGGTGTCTTCCAGTACGCCCGCGGGGG
CCCTTCTGCCCGAAGGACCGCTCGCGGCGAGCTTCTGCGGCGGCGACAAGCTCGCCCTCGTCCCAAGAAGTTCATGCGGGGCGCGCCCC
P F V P K R A L A A D F V G R Q E L P L L T N K W Y A G A P

CCGGTCTCGACCGCGGTCCGGAGCTGCCCGAACAGGGCGAGGAGCGGCCGGAAGGCGTCCCGGTGACCGCCACGGTCGTGG
GGCCAGAGCTGGCGCCACGCCCTCGAGGCCCTGGACGGGCTTGTCCTCCGGGCTTCGCGAGGCCAGCTCGGGGTCCAGCACCC
G T E V A T R L E P V Q G F L A L L R R F A D R D V G L D H

CGGGCCTCCTCAGCGGGGTGAAGGGCTGTTGCCCTAGCGCACGGCGAGCCGGACGAGGTGGCGGGCGGTCTGTTCCGGCCTCGTCGGGC
GCCCGAGGAGGTGCCCCCACTTCCCCGACAACGGGCATCGGTGCCGTCCGCTCGGCCTGCTCCACGCCCGCCAGCAAGGCCGAGGACGACCCG
R A E E L P T F P S N G Y R V A L R V L H R A T T G A E D P

FIG. 32FF

147/164

sugar.final gene b-1 Sequence

10	20	30	40	50	60	70	80	90
1234567890123456789012345678901234567890123456789012345678901234567890								

11610
GGCAGAGGCCCGCGGCGGAGGGTCTGGCCGACGGCGGTGGACCGCCGCCCCAGATCGGCTCCGGGGTGCGCGCAGCGTTCCGGCCGGG
CCGTGCTCCGGCGGCGCGCGCTCCAGACCGGCTGCCGACCTGGCGCGGGGGTCTAGCCGAGGCCCCACGCGCGTCCGCAAGCCGGCCCC
P V L G G A A L T Q G V A H V A A G L D A G P H A C R E A P

11700
GCGGTGGCGGAAGGCGGGCGGTCTCGGAGCGTCCAATCGTGGCGTGGATGTCTGGGGGCCGCGAGCGGGGGCGGGGCCGTGT
CGCCACCGCCTTTCCCGCCCCCGCCAGTAGCCCTCGCAGGTAGCACCCCGCACCTACAGACCCCCCGGCGCTGCCCGCCCCCGGCACA
A T A S L A P A T M R S R G I T P T S T Q P A A L P A P A T

NotI
▼

11790
CGCGGTGGCGCGGTCAAGTTCGCGCGCGGGTTCGCGCAGAGACGAGCGAGGTCCGGCGACCCGGCGGATGTCTGTCGCCCGATGGCGG
GCGCCACCGCGCGCAGTCAAGCGCGCGCGCGCTCTCTGCGTCTGCGTCCAGCGCTGGCGCGCTACAGCAGCAGCGGCTACCGCC
D R H R A T L E R R G R T A C L R L L D A V R R I D D G I A

PflMI
▼

11880
TGCCCGTCCGACGAGCAGCAGCGCGCGGCGGAGGCGTTCGGTGTGGCGCAGCGGGCGTCCCGCGGTACGGCTCCAGCTCGT
ACGGCCAGCGTCCCTGTCTGTCGCGCGCGCGCTCCGCAAGCCACACGCGTCCGCCCGCCGACGCCGACGGGGGCCATGCCGAGGTCGAGCA
T G T P L S L V R A A L R E T H P L P A H P Q G R Y P E L E

FIG. 32GG

148/164

sugar.finalgene b-1 Sequence

10 20 30 40 50 60 70 80 90
123456789012345678901234567890123456789012345678901234567890

▼
ApaLI

GGCAGCCCGGAGAGTAGGCGGGGTGTGACGCCCTTCGGCCTTCAGGACCTCCATGACGAGGTCCGGTGGATGCCGGTGGTGGCCT
CCGTCGGGCGGCTCTTCATCCGCGCCACACGTGCGGAAGCCGGAAGTCTTGAGGTACTGCTCAGCGCCACCTACGGCCACCCACCGGA
11970

H C G P S F Y A R T H V G E A K L V E M V L D R H I G T T A

▼
BsaAI

CGTCGATCTCGACGATCACGTACTGGTGGTTGTGAGCCCGTGGCGGTGCTGGCGACGAGGACGCCGGGAGGTCCGCGAGGTGCT
GCAGCTAGAGCTGCTAGTCATGACCAACCAACTCCGGCACCGCCAGCACCGCTGCTCTGGGCCCCCTCCAGGCGCTCCACGA
12060

E D I E V I V Y Q H N N L G H R D H D A V L V G P L D A L H

MluI
AflIII
▼

StyI
NcoI
▼

CGCGGTAGGCGGCGTGTGCGCCGGTTCGGTTCGATGACCTCGGAAACGCGTCGAGGAGGTGAGGCCCATGGCGGCGGCGCTCGC
GCGCCATCCGCGCACCAACGCGGCCAAGGCCAGCTACTGGAGCCCTTTGGCAGCTCCCTCCACTCCGGGTACCGCCGCCCGGAGCG
12150

E R Y A A H N R R N R D I V E P F A D L S T L G M A A A E

▼
BamHI

TCATCTTGGCGTTGGTCCCGCGGGGGTCCCGCGGGCAGGTGGAAGCCGAAGTTGTGAGGGCGCGGATCCGGGCGGCGAGGTGCG
AGTAGAACCGCAACAGGGCGGCGCGCCCGACGGGCGCGCTCCAGCTTCGGCTTCAACACCTCCCGCGCTAGGCCCGCGCTCCAGCC
12240

S M K A N T G G A P S G G P L D F G F N H L A R I R A A L D

FIG. 32HH

[illegible]

FIG. 32II

150/164

sugar.finalgene b-1 Sequence

10	20	30	40	50	60	70	80	90
1234567890123456789012345678901234567890123456789012345678901234567890								
BclI								
▼								
TGATCACTTCGCGGTGAGGCGGCGGTGCGGAGGAGCTGGAGCCCGCGCGTGGCGTTGCAGGTGGCCACGGCATGCCGGACCCCGG								12690
ACTAGTGAAGCGGCCACTCCGCGCGCGCACGCGCTCTCGACCTCGGCGCGGCACCGCAACGTCCACCGGTGCCGTACGGCCTGGGGCC								
M I V E G T L G A A H A L L Q L G A T A N C T A V A H R V G								
AlwNI								
▼								
CGAGCCCGCGACGCGCTCTCGAACTCGCGGACGAGCGGCGCGCGTGGACAGCCACTGGTGTGAGGGCCCCGGTCGAGCCGCTCGT								12780
GCTCGGGCCGCTGCGCGAGGAGCTTGAGCGCTGTGCTCGCCCGCGGGCAACCTGTGCGTGACCGACAGCTCCCGGGCCAGCTCGGCGAGCA								
A L G A V R E E F E R V L P G G N S L W Q S D L A R D L R E								
BsmI								
▼								
ACAGCTGGCGGGTTCGATGCGGTTGGGCGCGCCACGAGGAGCGGCTGGTCGAAAGCGGCGGGGCGCCCGAAGATGCCAGGTCGGATA								12870
TGTCGGACCGCGCCAGCTACGCCAACCCGCGGGGTGCTCTCGCCGACCCAGCTTTCGCGCGCGCCCGCGGCTTCTTACGCTCCAGCCTAT								
Y L R A R D I R N P R G V L L P Q D F A A P G G F A L D S								
TfII								
XmnI								
▼								
AGGCGCTTTTCACGGATGTTCCCTCGGGGCCACCGTCACGAAATGATTGCGCGGATCCGGGAATCCCGAACGAGGTGCGCGCGCTCCACCG								12960
TCCGCGAAAGTGCCTACAAAGGAGGCCCCGGTGGCAGTGCTTTACTAAGCGGCTAGGCCCTTAGGGCTTGCTCCACGCGCGCGGAGGTGC								
L A S K V								

FIG. 32JJ

151/164

sugar.finalgene b-1 Sequence

10	20	30	40	50	60	70	80	90
1234567890123456789012345678901234567890123456789012345678901234567890								
TGACGTACGACGAGATGGTCGATTGTGGTGGTTCGGGGGACTCTAATCCGCGCGGAACGGGACCCGACAAGAGCACGCTATGCGC								
ACTGCATGCTGCTCTACAGCTAACACACCAGCTAAAGCCCCCTGAGATTAGCGCGCCCTTGCCCTGGCTGTTCTCGTGCATACGGC								
13050								
TCTCGATGTGCTTCGGATCACATCCGCCCTCCGGGGTATTCCATCGGGGGCCCGAATGTGATGATCCTTGACAGGATCCGGGAATCAGCCG								
AGAGCTACACGAAGCCTAGTGTAGCGGAGGCCCCCATAAGGTAGCCCGCGGCTTACACTACTAGGAACCTGCTCCTAGGCCCTTAGTTCGGC								
13140								
AGCCGCGGAGGGCGCGCTCCGCGGAAGAGTACGTGTGAGAAAGTCCCGTTCCTCTTCCCGTTTCCGTTCCGCTTCCGGCCCCCG								
TCGGCGGCCCTCCCGGCCCGCGGAGCGCCTTCTCATGCACACTCTTCAGGGCAAGGAGAAAGGCAAGGCAAGGCCGAAGGCCGGGCC								
13230								
TCTGGAGTTCTCCGTGCGCCGTACCAGCAGGGAACGACCGCTTCTCCCCGGTACTCGACCTCGGGGCCCTGGGGCAGGATTTCCGGCGC								
AGACCTCAAGAGGCACGCGGCATGGGTCCCTTGCTGGCGAAGAGGGGCCCATGAGCTGGAGCCCGGGACCCCGTCTTAAGCGCGC								
13320								
V R R T Q Q G T T A S P P V L D L G A L G Q D F A A								

FIG. 32KK

sugar.finalgene b-1 Sequence

[illegible][illegible]

FIG. 32LL